

Hybrid crops that never lose their vigor p. 684

The emergence of kissing in ancient societies p. 688

A wearable device restores thermal sensation in amputees p. 731

Science

$15
19 MAY 2023
science.org
AAAS

SHRINKING LAKES

Global water storage in large lakes is decreasing

pp. 693 & 743

VISIT…

LANZAROTE Caliente.com

EDITORIAL

Don't let the Court disrupt science

Adam F. Falk
is president of the Alfred P. Sloan Foundation, New York, NY, USA.
falk@sloan.org

Lorelle L. Espinosa
is a program director at the Alfred P. Sloan Foundation, New York, NY, USA.
espinosa@sloan.org

It is widely expected that by this summer, the United States Supreme Court will overturn long-standing precedents allowing the consideration of race as one factor among many in university admissions. The current legal regime goes back to the Court's decision (*Regents of University of California* v. *Bakke*) in 1978 that banned racial quotas while allowing consideration of race for the purpose of creating a diverse educational environment. Although the law has evolved since then, almost all universities have relied on the *Bakke* framework to support their strategies to educate a diverse citizenry. If the Court upends those practices, the implications for the scientific enterprise will be far-reaching. It is essential that the process of science continue to become more diverse, equitable, and inclusive. Studies show that the best science is done when teams are diverse. Moreover, the very questions scientists address can change substantially when those scientists come from a variety of racial, ethnic, and other backgrounds.

The upcoming decisions are likely to be disruptive in ways that go far beyond admission practices. Many observers expect the Court to issue a broad ruling that affects student benefits more generally. For example, the Court could prohibit the consideration of race in conferral of scholarships or participation in mentoring, enrichment, or bridge programs—all approaches known to be successful in broadening participation in science. And even if the ruling is a narrow one, the prospect of subsequent lawsuits and confusion about new restrictions likely will have a chilling effect on existing efforts to advance diversity in science. Race-neutral considerations such as socioeconomic status, although important, are insufficient to redress racial inequities. The impact of racism goes well beyond economics.

> "The upcoming decisions are likely to be disruptive in ways that go far beyond admission practices."

In response to this new legal environment, it will be critical to continue, within the law, to uphold commitments to diversity in science. The Diversity and the Law project of AAAS (the American Association for the Advancement of Science, the publisher of *Science*) has developed useful legal guidance, which will be updated in the wake of the rulings. But the scientific community must acknowledge that the programs under pressure from the Court are precisely those designed to help individual students access higher education and succeed in an unwelcoming, sometimes hostile, scientific culture. Moreover, these programs are limiting, because a focus solely on representational diversity does nothing to remedy the environments that individuals experience, including in college classrooms and research laboratories.

The upcoming ruling should be a clarion call to improve the scientific environment itself. The need for systemic cultural change was emphasized earlier this year by the US National Academies report, "Advancing Antiracism, Diversity, Equity, and Inclusion in STEMM Organizations: Beyond Broadening Participation." Its message is that unless the practice of science becomes more equitable and inclusive, efforts to bring more underrepresented scientists into the field are doomed to founder.

Fortunately, strategies for advancing cultural change need not be affected by the considerations currently before the Court, which are concerned largely with the conferral of individual benefits. What is more, a broad range of actors can contribute meaningfully to this effort, as many already are. Scientific societies can improve professional norms by providing robust resources and tools to their members. Faculties can adopt inclusive classroom pedagogy and create cultures of effective mentoring through participation in communities of learning and practice. Universities can redesign policies and practices to remove barriers to student retention, such as attending to students' needs beyond academics. They can establish partnerships with Minority Serving Institutions to create pathways into graduate programs. And although there are universities where the current political climate excludes diversity as an explicit stated goal of such initiatives, measures such as these benefit all students and should be framed as such.

Reforming the culture of science to be more equitable and inclusive is a key priority of the Alfred P. Sloan Foundation and many of our peers. We are making deep investments in graduate education and the professoriate, seeking institutional partners who are already committed to this transformation. The overwhelming response we have seen shows that many scientists are eager for an environment that is more diverse, equitable, and inclusive.

With an unshakable commitment and working together, we can assure that the upcoming Court decisions, as disruptive as they will be, will not disrupt the progress of science toward a better future.

–Adam F. Falk and Lorelle L. Espinosa

10.1126/science.adi7054

NEWS

IN BRIEF

Edited by **Jeffrey Brainard**



PLANETARY SCIENCE

Mars rover spots signs of big river

NASA's Perseverance rover has snapped images indicating a river on Mars once flowed fast and furiously, scientists said last week. The waterway was part of a network that flowed into Jezero crater, which the rover has been exploring since landing in February 2021. The new clues include rows of rocks that may be remnants of sandbars or shifting banks, in an area dubbed Skrinkle Haven (mosaic image, right). A river may have also formed a nearby hill, Pinestand, bearing sedimentary layers up to 20 meters thick, much thicker than similar deposits on Earth. Studying these features could shed light on the past climate of Mars and whether it could have been habitable for microbial life.

WHO cancels mpox emergency

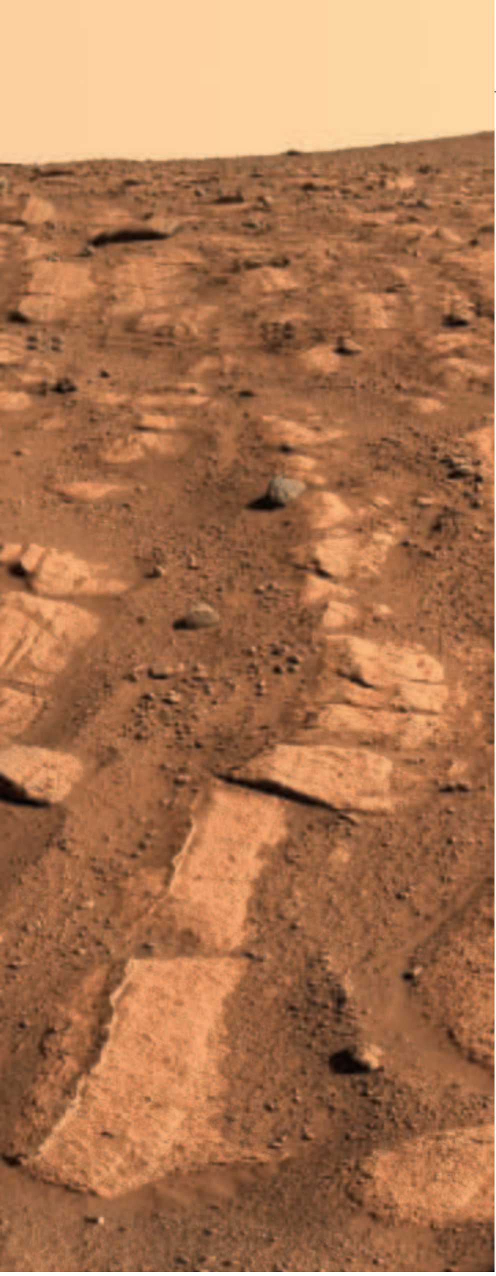
INFECTIOUS DISEASES | The World Health Organization (WHO) last week ended the global health emergency for mpox, 10 months after declaring it, as cases have fallen dramatically in Europe and the Americas. Caused by the monkeypox virus (MPXV), the disease is endemic in Africa, but a new strain surged elsewhere last year, spurring WHO to invoke a Public Health Emergency of International Concern. Nearly 90% fewer mpox cases were reported to WHO in the past 3 months compared with the preceding 3 months. In total, WHO has received reports of more than 87,000 mpox cases and 140 deaths from 111 countries. MPXV is spread by intimate personal contact. In the current outbreak, men who have sex with men have been predominantly affected. Mpox commonly causes a painful rash and flulike symptoms but can kill immunocompromised people. Natural immunity and changes in behavior helped slow the outbreak, but WHO said progress was hampered by a scarcity of vaccines and an absence of international funding.

Biden nominates NIH director

LEADERSHIP | It's official: President Joe Biden wants cancer researcher Monica Bertagnolli to be the next director of the U.S. National Institutes of Health (NIH). On 15 May Biden confirmed weeks of media speculation by nominating Bertagnolli to succeed Francis Collins, who stepped down in December 2021. Bertagnolli has spent the past 8 months as head of NIH's largest unit, the National Cancer Institute (NCI), and has championed Biden's signature Cancer Moonshot initiative, designed to cut the U.S. death rate from cancer 50% by 2050. Bertagnolli faces a grilling in the U.S. Senate on issues such as NIH's funding for a Chinese lab that some believe was the cause of the pandemic and how the agency can help lower drug prices. But the Democrat-controlled body is expected to confirm her.

Spider ecologist's fraud confirmed

RESEARCH MISCONDUCT | McMaster University last week released results from its investigation into former faculty member Jonathan Pruitt, finding that the internationally acclaimed behavioral ecologist engaged in fabrication and falsification of data reported in eight papers. Starting in 2020, co-authors brought allegations of data manipulation. McMaster placed Pruitt on leave in 2021, and they



resigned in 2022 in a settlement. To date, 15 of Pruitt's papers on social behavior in spiders have been retracted, work that had earned them a prestigious Canada 150 Research Chair title. Pruitt declined *Science*'s request for comment, but said, "I'll be available to chat in the fall when my first book comes out."

EU science leader steps down

LEADERSHIP | The European Commission's science chief has left her post 1 year before the end of her term, as decisions loom over some of the EU's science priorities. On 10 May, Mariya Gabriel was granted unpaid leave from her EU job as commissioner overseeing multiple portfolios—innovation, research, culture, education, and youth—to help form a government in Bulgaria, her home country. She was tapped by the head of Bulgaria's conservative party to try to break a political deadlock; if she succeeds, Gabriel will become prime minister. Her absence comes as the European Commission is developing a strategic plan for the second part of the current €95 billion research program, Horizon Europe, which would start in 2025 and last 3 years. For now, Gabriel's portfolios will be covered by two Commission vice presidents.

U.S. eases blood screening for HIV

PUBLIC HEALTH | In a move aimed at making it easier for men who have sex with men (MSM) to give blood, the U.S. Food and Drug Administration (FDA) last week issued revised recommendations for screening potential donors for HIV. Previously, the guidelines called for asking people about their sexual orientation, a practice some called discriminatory. Now, screening will ask all volunteers, regardless of gender identity, whether they recently had unprotected, risky sex. Monogamous MSM will no longer be automatically excluded. Because the nucleic acid tests used to screen blood for HIV may not detect recent infections, FDA now recommends that blood banks defer all prospective donors, regardless of gender identity, who report having a new sexual partner or more than one sexual partner in the past 3 months and who've had anal sex in that span. The agency says the new recommendations are based on data from countries with similar policies, including Canada and the United Kingdom.

State yanks Kinsey center funds

SEX RESEARCH | A spending bill signed on 4 May by Indiana's governor bars state funding for the famed research center on human sexuality founded by Alfred Kinsey in 1947. Indiana University Bloomington, which houses the Kinsey Institute, has drawn on the university's overall state appropriation to help support the center and its 11 research faculty members; the institute will continue to receive the majority of its funding from external grants and donations. But the spending measure "singling out a specific research institute sets a troubling precedent with implications that could limit the ability of public colleges and universities to pursue research and scholarship that benefits people and improves lives," said Pamela Whitten, the university's president, last month. The funding prohibition was introduced as an amendment to the bill by state Representative Lorissa Sweet, a Republican who alleged the institute has enabled sexual abuse. State Representative Matt Pierce, a Democrat whose district includes Bloomington, said Sweet's claims were based on "old, unproven allegations of conspiracies" that have reappeared as "warmed-over internet memes."

Australia to set up CDC

PUBLIC HEALTH | Australia is establishing a Centre for Disease Control (CDC), joining the majority of advanced countries in having a national health agency. Australia's states and territories will retain responsibility for public health; the new CDC is meant to provide nationwide surveillance of disease outbreaks and help improve coordination, "something that was sorely lacking during COVID," says Ben Marais, an infectious disease specialist at the University of Sydney. The 2023–24 federal budget announced last week includes $61 million over 2 years to get the center up and running.

Extrasolar radiation belt spotted

ASTRONOMY | Researchers have spied the first radiation belt outside the Solar System. These particle racetracks—rings of high-speed electrons and other charged particles—circle planets with strong magnetic fields, including Earth, Jupiter, and Saturn, but it was not known whether they existed in other planetary systems. A team reported in *Nature* this week that they found one around a Jupiter-size brown dwarf—a gas giant too small to have been a hydrogen-fusion star—about 20 light-years away. Electrons whipping around the giant ring, 18 Jupiter diameters across, emitted radio waves detected by an array of 39 radio telescopes spanning the globe.

BY THE NUMBERS

$451 billion

Upper estimate of the cost of unequal health risks and access to health care borne by U.S. racial and ethnic minorities in 2018. Costs included premature death, lost work productivity, and excess spending for medical care, amounting to 2.2% of gross domestic product. Among all Americans including white people, adults without a 4-year college degree had costs up to $978 billion. (*JAMA*)

IN DEPTH



Researchers found abundant human DNA in the waters of the Avoca River in Ireland.

BIOETHICS

Human genomic 'bycatch' threatens privacy

Environmental and microbe studies net human DNA that can reveal health, identity

By **Gretchen Vogel**

Everywhere they go, humans leave stray DNA. Police have used genetic sequences retrieved from cigarette butts and coffee cups to identify suspects; archaeologists have sifted DNA from cave dirt to identify ancient humans. But for scientists aiming to capture genetic information not about people, but about animals, plants, and microbes, the ubiquity of human DNA and the ability of even partial sequences to reveal information most people would want to keep private is a growing problem, researchers from two disparate fields warn this week. Both groups are calling for safeguards to prevent misuse of such human genomic "bycatch."

Genetic sequences recovered from water, soil, and even air can reveal plant and animal diversity, identify pathogens, and trace past environments, sparking a boom in studies of this environmental DNA (eDNA). But the samples can also contain significant amounts of human genes, researchers report this week in *Nature Ecology & Evolution*. In some cases, the DNA traces were enough to determine the sex and likely ancestry of the people who shed them, raising ethical alarms.

Similarly, scientists have for decades analyzed the genetic information in fecal matter to reveal the microbes in people's intestines—the gut microbiome, which plays dramatic roles in human health and development. Most researchers have assumed that the traces of degraded human DNA found in stool don't contain significant genetic information about the sample provider. But enough is present, according to an analysis published this week in *Nature Microbiology*, to potentially identify the donor's sex, likely ancestry, certain disease risks, and, when linked to other databases, even their full identity. Computer programs commonly used to filter out human genetic sequences from the microbiome data did not eliminate the problem, the researchers found.

"I see this as a major ethics problem for the whole field," says microbiome pioneer Rob Knight of the University of California, San Diego (UCSD), who wrote an accompanying commentary. "We will have to completely rethink how we communicate to research subjects about the privacy risks of participating in microbiome research."

The genomic bycatch issue could have far-reaching consequences. Knight says his group may need to take down all of the genetic sequences it has posted to public databases to remove more of the human DNA. It will also complicate data sharing among eDNA researchers and could mean that ecologists and environmental scientists, who are not used to getting permits to study people, will need another set of ethics approvals. eDNA or microbiome data from groups, such as Indigenous peoples, who have long been concerned about the unintended consequences of genetic data, could be especially sensitive. "We're moving in a direction and at a pace that we really need to recalibrate" the ethics and legal rules surrounding such data, says Keolu Fox, a genomic scientist at UCSD who is Kānaka Maoli (Native Hawaiian) and has long advocated for Indigenous rights in genetic research.

Jessica Farrell, David Duffy, and their colleagues at the University of Florida were using eDNA to probe herpes virus infections that cause tumors in sea turtles when they started to worry about human bycatch. They applied a particularly powerful sequencing method to samples of sand from nesting sites and water from tidal estuaries and tanks at the university's Sea Turtle Hospital. They found the turtle virus and turtle DNA, but they also found long stretches of human DNA, intact enough that sequencing programs could easily recognize X and Y chromosomes.

Startled by the abundant human DNA in their samples, they received ethics board

permission to look for human DNA in a variety of other environments. They found it in samples from the Avoca River in Ireland; seawater from near the coast of St. Augustine, Florida; sand from a footprint; and air in a room where people were working. In many of the samples, the genetic traces were enough to identify the sex and likely ancestry of the person who left them, and some revealed genetic variants associated with disease risk.

In contrast, human geneticists Yukinori Okada and Yoshihiko Tomofuji of Osaka University and their colleagues set out from the start to see whether stray human sequences in their microbiome data posed a privacy problem. They had sequenced both the genomes and fecal microbiomes of 343 people to study how a person's genes might correlate with and possibly influence their gut flora, and they wanted to be sure that if they shared any of their data, the donors would not be identifiable. What they found was not reassuring. For 97% of the samples, the microbiome data contained enough human genetic information to correctly predict the donor's sex.

Almost all the donors could also be reidentified from single nucleotide polymorphisms (SNPs) in the fragments of human DNA present in the microbiome data. SNPs are specific points in the genome where the DNA sequence typically varies between people; they are the basis of DNA fingerprinting, and certain SNPs can also determine susceptibility to some diseases. The microbiome data didn't include enough information for a standard DNA fingerprint match, but by applying advanced statistical methods to the combinations of SNPs in each sample, the researchers matched 320 of the 343 fecal samples to the correct donor. Even after trying various computer programs to screen for and filter out human DNA sequences, the group could still match up to 39 of the samples with the correct donor.

They could also confirm the East Asian ancestry of all but six of their 343 donors. When they used the same techniques on publicly available gut microbiome data sets from Europe, South Asia, and East Asia and assumed the donors' ancestry matched the data's region, they could correctly predict ancestry between 80% and 92% of the time. As a final step, the scientists did additional "ultradeep shotgun" sequencing on five of the fecal samples. They identified genotypes associated with inflammatory bowel diseases, type 2 diabetes and other, even rarer conditions, information that is usually considered highly sensitive and private.

Knight says the analysis shows that the current methods microbiome researchers use to filter out human DNA and anonymize samples simply don't work well enough. Researchers also need to re-evaluate how widely microbiome-derived sequences can be shared, he adds. "We are currently looking at withdrawing all the human metagenomic data sets we ever deposited, so we can redeposit only the sequences that positively match a microbe," he says.

The power to extract personal data from eDNA and microbiome samples will continue to increase, both groups of authors warn. That raises concerns about misuse by police or other government agencies, collection by commercial companies, or even mass genetic surveillance, says Natalie Ram, a law and bioethics scholar at the University of Maryland Francis King Carey School of Law. In the United States, she says, researchers and funding agencies should make greater use of federal Certificates of Confidentiality. They prohibit the disclosure of "identifiable, sensitive research information" to anyone not connected with a study, such as law enforcement, without the subject's consent.

> **"I see this as a major ethics problem for the whole field."**
> **Rob Knight**, University of California, San Diego

The National Institutes of Health automatically issues such certificates for federally funded health-related research, but could expand the range of studies that are covered, Ram says. In Europe, says Ewan Birney, a bioinformaticist and deputy director general of the European Molecular Biology Laboratory, existing data protection laws should help protect against misuse of human genomic bycatch while still allowing research to proceed. "It would be a bad thing for the world if we were not able to share eDNA samples," he says.

Yves Moreau of KU Leuven, who studies both artificial intelligence and genetics and has warned that China is using mass collections of human DNA to help suppress minorities, agrees that any limits on research need to be carefully weighed with the potential benefits. At the same time, he says, "it is important to put this on everybody's radar so that emerging risks and potential abuses can be identified early."

Fox concurs. "Which companies and governments are going to pay and license to have poop-based surveillance technology?" he asks. "Imputing people's identity based on their poop is compelling and interesting, for a number of reasons, and most of them are all the wrong reasons." ■

With reporting by Elizabeth Pennisi and Rodrigo Pérez Ortega.

BIOSECURITY

Benchtop DNA synthesis raises alarms

Report calls for safeguards against misuse for bioterror

By **Robert F. Service**

DIY has come to DNA synthesis: Benchtop machines can print DNA in the lab, eliminating the need for scientists to order sequences from a synthetic biology company. But this technology brings new risks by circumventing how companies now screen for would-be bioterrorists, a report last week by a Washington, D.C., think tank warns.

The current screening system, which is voluntary, "could be upended by benchtop DNA synthesis," says report co-author Jaime Yassif, vice president for global biological policy and programs at the Nuclear Threat Initiative. "Governments, industry, and the broader scientific community need to put stronger safeguards in place," she says.

The ability to synthesize DNA is critical to basic research and biotechnology. Roughly 100 companies now print DNA sequences to order. To keep malign actors from synthesizing DNA to make a powerful toxin or bioweapon, the U.S. government released voluntary guidelines in 2010 recommending that companies vet customers and screen ordered sequences against known dangers. Most companies abide by the standards, but attempts to mandate them "have been really, really slow," says Elizabeth Cameron, a biosafety expert at Brown University.

Benchtop DNA printers heighten those concerns. Current machines can synthesize stretches of about 200 base pairs, but the report notes that over the next 2 to 5 years, the sequence lengths could increase to as many as 7000 base pairs. That's the size of the smallest viruses, and big enough to splice together large pathogen genomes. The report recommends safeguards such as requiring makers of benchtop synthesizers to check that their customers are legitimate researchers and build in software that blocks the synthesis of risky sequences. Mike Daniels of Evonetix, a DNA synthesis devicemaker, says he supports such measures. "We need a strong and clear baseline of minimally acceptable standards." ■

permission to look for human DNA in a variety of other environments. They found it in samples from the Avoca River in Ireland; seawater from near the coast of St. Augustine, Florida; sand from a footprint; and air in a room where people were working. In many of the samples, the genetic traces were enough to identify the sex and likely ancestry of the person who left them, and some revealed genetic variants associated with disease risk.

In contrast, human geneticists Yukinori Okada and Yoshihiko Tomofuji of Osaka University and their colleagues set out from the start to see whether stray human sequences in their microbiome data posed a privacy problem. They had sequenced both the genomes and fecal microbiomes of 343 people to study how a person's genes might correlate with and possibly influence their gut flora, and they wanted to be sure that if they shared any of their data, the donors would not be identifiable. What they found was not reassuring. For 97% of the samples, the microbiome data contained enough human genetic information to correctly predict the donor's sex.

Almost all the donors could also be reidentified from single nucleotide polymorphisms (SNPs) in the fragments of human DNA present in the microbiome data. SNPs are specific points in the genome where the DNA sequence typically varies between people; they are the basis of DNA fingerprinting, and certain SNPs can also determine susceptibility to some diseases. The microbiome data didn't include enough information for a standard DNA fingerprint match, but by applying advanced statistical methods to the combinations of SNPs in each sample, the researchers matched 320 of the 343 fecal samples to the correct donor. Even after trying various computer programs to screen for and filter out human DNA sequences, the group could still match up to 39 of the samples with the correct donor.

They could also confirm the East Asian ancestry of all but six of their 343 donors. When they used the same techniques on publicly available gut microbiome data sets from Europe, South Asia, and East Asia and assumed the donors' ancestry matched the data's region, they could correctly predict ancestry between 80% and 92% of the time. As a final step, the scientists did additional "ultradeep shotgun" sequencing on five of the fecal samples. They identified genotypes associated with inflammatory bowel diseases, type 2 diabetes and other, even rarer conditions, information that is usually considered highly sensitive and private.

Knight says the analysis shows that the current methods microbiome researchers use to filter out human DNA and anonymize samples simply don't work well enough. Researchers also need to re-evaluate how widely microbiome-derived sequences can be shared, he adds. "We are currently looking at withdrawing all the human metagenomic data sets we ever deposited, so we can redeposit only the sequences that positively match a microbe," he says.

The power to extract personal data from eDNA and microbiome samples will continue to increase, both groups of authors warn. That raises concerns about misuse by police or other government agencies, collection by commercial companies, or even mass genetic surveillance, says Natalie Ram, a law and bioethics scholar at the University of Maryland Francis King Carey School of Law. In the United States, she says, researchers and funding agencies should make greater use of federal Certificates of Confidentiality. They prohibit the disclosure of "identifiable, sensitive research information" to anyone not connected with a study, such as law enforcement, without the subject's consent.

> **"I see this as a major ethics problem for the whole field."**
> **Rob Knight**, University of California, San Diego

The National Institutes of Health automatically issues such certificates for federally funded health-related research, but could expand the range of studies that are covered, Ram says. In Europe, says Ewan Birney, a bioinformaticist and deputy director general of the European Molecular Biology Laboratory, existing data protection laws should help protect against misuse of human genomic bycatch while still allowing research to proceed. "It would be a bad thing for the world if we were not able to share eDNA samples," he says.

Yves Moreau of KU Leuven, who studies both artificial intelligence and genetics and has warned that China is using mass collections of human DNA to help suppress minorities, agrees that any limits on research need to be carefully weighed with the potential benefits. At the same time, he says, "it is important to put this on everybody's radar so that emerging risks and potential abuses can be identified early."

Fox concurs. "Which companies and governments are going to pay and license to have poop-based surveillance technology?" he asks. "Imputing people's identity based on their poop is compelling and interesting, for a number of reasons, and most of them are all the wrong reasons." ■

With reporting by Elizabeth Pennisi and Rodrigo Pérez Ortega.

BIOSECURITY

Benchtop DNA synthesis raises alarms

Report calls for safeguards against misuse for bioterror

By **Robert F. Service**

DIY has come to DNA synthesis: Benchtop machines can print DNA in the lab, eliminating the need for scientists to order sequences from a synthetic biology company. But this technology brings new risks by circumventing how companies now screen for would-be bioterrorists, a report last week by a Washington, D.C., think tank warns.

The current screening system, which is voluntary, "could be upended by benchtop DNA synthesis," says report co-author Jaime Yassif, vice president for global biological policy and programs at the Nuclear Threat Initiative. "Governments, industry, and the broader scientific community need to put stronger safeguards in place," she says.

The ability to synthesize DNA is critical to basic research and biotechnology. Roughly 100 companies now print DNA sequences to order. To keep malign actors from synthesizing DNA to make a powerful toxin or bioweapon, the U.S. government released voluntary guidelines in 2010 recommending that companies vet customers and screen ordered sequences against known dangers. Most companies abide by the standards, but attempts to mandate them "have been really, really slow," says Elizabeth Cameron, a biosafety expert at Brown University.

Benchtop DNA printers heighten those concerns. Current machines can synthesize stretches of about 200 base pairs, but the report notes that over the next 2 to 5 years, the sequence lengths could increase to as many as 7000 base pairs. That's the size of the smallest viruses, and big enough to splice together large pathogen genomes. The report recommends safeguards such as requiring makers of benchtop synthesizers to check that their customers are legitimate researchers and build in software that blocks the synthesis of risky sequences. Mike Daniels of Evonetix, a DNA synthesis devicemaker, says he supports such measures. "We need a strong and clear baseline of minimally acceptable standards." ■



Emissions from petrochemical plants are degrading petroglyphs on Australia's Burrup Peninsula, researchers say.

ARCHAEOLOGY

Australia's rock art threatened by development

Petrochemical pollution is eroding world-famous Murujuga petroglyphs, scientists warn

By **Dennis Normile**

The Murujuga rock art, the world's largest and oldest collection of petroglyphs, has survived 40,000 years or so out in the open on a finger of land extending from Australia's northwestern coast. "Murujuga is the most important rock art site in the world," says Michel Lorblanchet, an archaeologist at CNRS, the French national research agency, who is noted for his work on European cave art.

It may also be the most threatened. Scientists and the First Nations custodians of the site are warning that acidic emissions from a nearby petrochemical complex are etching away images densely spread across the 30-by-6-kilometer Burrup Peninsula and on nearby islands. And the pollution is about to get worse. Last month work began on a new fertilizer plant, and a new natural gas processing facility is in planning.

If realized, the expanded complex "would be the most polluting project ever to be developed in Australia," asserts the Conservation Council of Western Australia. The additional emissions could make it difficult for Australia to meet its commitment to reducing greenhouse gas emissions, says Bill Hare, a climate scientist with Climate Analytics, a policy institute. As for the rock art, "We will start to see degradation very clearly within people's lifetimes," predicts Benjamin Smith, a rock art specialist at the University of Western Australia.

The Murujuga rocks look like random jumbles of blocks with a reddish brown patina slowly built up by mineralization. Some 40,000 years ago, Aboriginal artists began pecking and scratching through the patina to expose the grayish rock beneath, ultimately creating more than 1 million images. "When these rock art images were made, they would have been very visible," Smith says.

The rock art provides an archaeological record of traditional use of the area over thousands of years. The petroglyphs include some of the world's earliest known depictions of human faces, complex geometric designs, and images of marine and terrestrial fauna, including extinct animals such as the fat-tailed kangaroo and the thylacine, or Tasmanian tiger. They illustrate how the ecosystem changed through rising and falling sea levels. The artistry ended in 1868, when European settlers decimated the local Yaburara People in what is known as the Flying Foam Massacre.

Today, the rock sites hold a spiritual significance for First Nations peoples that many liken to what European cathedrals once meant for Christians. But the rock art is even more vital and important because Australia's First Peoples never developed a written language. Murujuga "is our parliament, it is a gathering place where we have our politics, our governance, our laws, and our protocols," says Josie Alec, a member of the Kuruma and Mardudhunera peoples and a Murujuga traditional custodian. Looking after Ngurra, or Earth, "is a way of making sure humanity's future generations have a safe place to live," adds Raelene Cooper, a member of the Mardudhunera People who also serves as a custodian.

In the 1960s, when officials were looking to build a cargo port on the northwestern coast to ship iron ore from inland mines, they knew little about the Murujuga rock art and did not consult with local First Nations peoples. Antiquities experts recommended the Burrup Peninsula to avoid damaging other known Aboriginal sites. Later, as the significance of the Murujuga rock art was recognized, people realized the

port siting was a "terrible mistake," Smith says. The port attracted other industry, including a large sea salt–evaporation operation. The discovery of offshore gas fields spurred the development of gas processing facilities and a mammoth fertilizer plant. The peninsula now hosts Australia's largest petrochemical complex.

Construction projects destroyed an estimated 10,000 petroglyphs before the rock art began to receive protection. In 2013, more than 40% of the peninsula was incorporated into Murujuga National Park, which is jointly managed by the Murujuga Aboriginal Corporation and the Western Australia state government.

That move hasn't protected the art from air pollution, researchers say. Satellite observations show the natural gas plants create clouds of nitrogen oxides. Smith and colleagues have shown that dust and rain drop oxides onto rock surfaces, and lab experiments indicate the oxides form acids that erode the patina. Reducing industrial emissions is "essential if damage to the rock art is to be limited and this iconic cultural place is to remain largely intact for future generations," the researchers wrote last year in *Conservation and Management of Archaeological Sites*.

In 2021, Alec and Cooper formed the group Save Our Songlines, which is demanding a halt to the new fertilizer and natural gas projects. Allowing construction to proceed will "harm our world and humanity with emissions," Cooper says.

So far, however, government officials have only committed to review the potential impact of emissions under a provision of a federal Aboriginal and Torres Strait Islander Heritage Protection Act. The review could stop the projects, but opponents say that is unlikely.

Woodside Energy, the company behind the existing and proposed natural gas plants, claims that the emissions do not pose a threat. "Peer-reviewed research has not demonstrated any impacts on Burrup rock art from emissions associated with Woodside's operations," reads a statement on its website. But in a recent study that compared past and current photographs of 26 petroglyphs, Smith's group found that half were damaged, two substantially. Those closest to the petrochemical complex suffered the greatest degradation.

Smith hopes worldwide concern for the rock art will put pressure on industry to "get its act together by putting [emissions] scrubbers in those stacks," and on the government to "get its act together by putting new industry somewhere else." Action is urgent, he says. "You can't remake this rock. Once it's gone, it's gone forever." ■

PHOTO: ELLA MARU STUDIO/SCIENCE SOURCE

CANCER BIOLOGY

New partners reinvigorate a once-touted cancer treatment

Decades after faltering, angiogenesis inhibitors show new promise when paired with checkpoint drugs

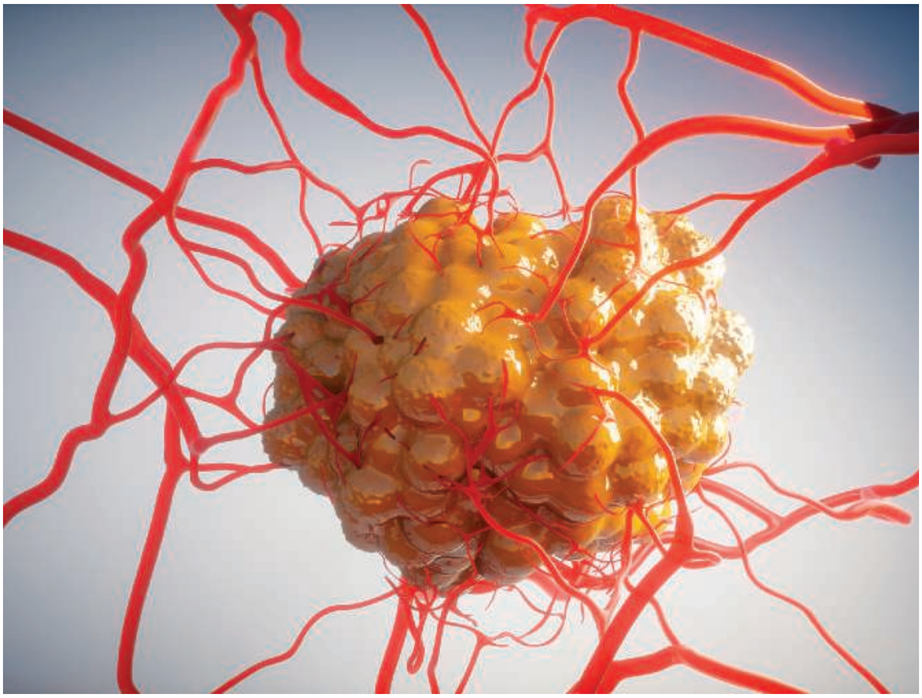
By **Mitch Leslie**

Patients with a type of liver cancer known as hepatocellular carcinoma (HCC) often face an anxious wait after their tumor is removed. In up to half of these people, the cancer will return within 2 years after surgery or a treatment that destroys their tumors with heat. Researchers haven't identified any therapies that can stop it from coming back.

That could change thanks to a class of drugs once seen as a failed revolution in cancer therapy. Angiogenesis inhibitors, which throttle tumors by cutting off their blood supply, never measured up to expectations when they reached clinical trials more than 2 decades ago. But they are now getting a boost from a newer set of drugs with a more impressive track record: checkpoint inhibitors, which unleash the immune system's T cells to attack tumors. At the American Association for Cancer Research (AACR) meeting last month, researchers presented evidence that the combination of an angiogenesis inhibitor and a checkpoint inhibitor delayed recurrence in patients with HCC, a first for this cancer type.

If the duo receives approval from the U.S. Food and Drug Administration (FDA), it will be the eighth pairing of the drug types OK'd in the past 4 years. More than 200 clinical trials are now testing this approach in various kinds of cancers, spurred by evidence that angiogenesis inhibitors help cancer-fighting T cells find their way deep into tumors. "There's a very clear rationale" for combining these two classes of drugs, says Robert Kerbel, a tumor immunologist at the University of Toronto.

This resurgence for angiogenesis inhibitors comes after astronomical expectations for the drugs were dashed in the late 1990s. Their premise was compelling. Tumors, hungry for oxygen and nutrients, would have to stimulate new blood vessels to sprout and grow to them; by preventing these vessels from forming, angiogenesis inhibitors should starve tumors and curb



Drugs that prevent new blood vessel growth may make tumors more vulnerable to checkpoint inhibitors.

port siting was a "terrible mistake," Smith says. The port attracted other industry, including a large sea salt–evaporation operation. The discovery of offshore gas fields spurred the development of gas processing facilities and a mammoth fertilizer plant. The peninsula now hosts Australia's largest petrochemical complex.

Construction projects destroyed an estimated 10,000 petroglyphs before the rock art began to receive protection. In 2013, more than 40% of the peninsula was incorporated into Murujuga National Park, which is jointly managed by the Murujuga Aboriginal Corporation and the Western Australia state government.

That move hasn't protected the art from air pollution, researchers say. Satellite observations show the natural gas plants create clouds of nitrogen oxides. Smith and colleagues have shown that dust and rain drop oxides onto rock surfaces, and lab experiments indicate the oxides form acids that erode the patina. Reducing industrial emissions is "essential if damage to the rock art is to be limited and this iconic cultural place is to remain largely intact for future generations," the researchers wrote last year in *Conservation and Management of Archaeological Sites*.

In 2021, Alec and Cooper formed the group Save Our Songlines, which is demanding a halt to the new fertilizer and natural gas projects. Allowing construction to proceed will "harm our world and humanity with emissions," Cooper says.

So far, however, government officials have only committed to review the potential impact of emissions under a provision of a federal Aboriginal and Torres Strait Islander Heritage Protection Act. The review could stop the projects, but opponents say that is unlikely.

Woodside Energy, the company behind the existing and proposed natural gas plants, claims that the emissions do not pose a threat. "Peer-reviewed research has not demonstrated any impacts on Burrup rock art from emissions associated with Woodside's operations," reads a statement on its website. But in a recent study that compared past and current photographs of 26 petroglyphs, Smith's group found that half were damaged, two substantially. Those closest to the petrochemical complex suffered the greatest degradation.

Smith hopes worldwide concern for the rock art will put pressure on industry to "get its act together by putting [emissions] scrubbers in those stacks," and on the government to "get its act together by putting new industry somewhere else." Action is urgent, he says. "You can't remake this rock. Once it's gone, it's gone forever." ■

CANCER BIOLOGY

New partners reinvigorate a once-touted cancer treatment

Decades after faltering, angiogenesis inhibitors show new promise when paired with checkpoint drugs

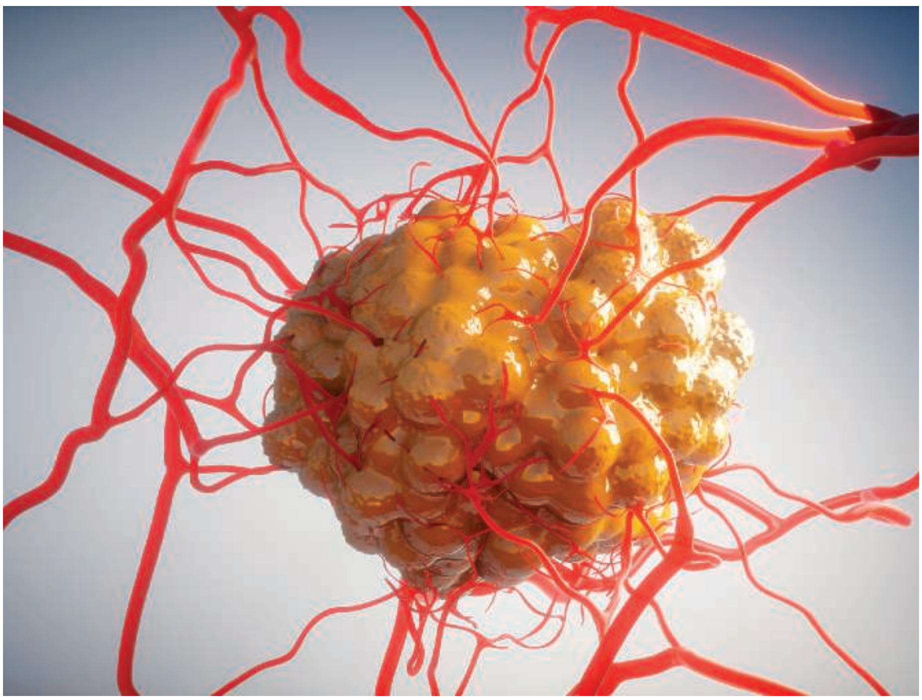
By **Mitch Leslie**

Patients with a type of liver cancer known as hepatocellular carcinoma (HCC) often face an anxious wait after their tumor is removed. In up to half of these people, the cancer will return within 2 years after surgery or a treatment that destroys their tumors with heat. Researchers haven't identified any therapies that can stop it from coming back.

That could change thanks to a class of drugs once seen as a failed revolution in cancer therapy. Angiogenesis inhibitors, which throttle tumors by cutting off their blood supply, never measured up to expectations when they reached clinical trials more than 2 decades ago. But they are now getting a boost from a newer set of drugs with a more impressive track record: checkpoint inhibitors, which unleash the immune system's T cells to attack tumors. At the American Association for Cancer Research (AACR) meeting last month, researchers presented evidence that the combination of an angiogenesis inhibitor and a checkpoint inhibitor delayed recurrence in patients with HCC, a first for this cancer type.

If the duo receives approval from the U.S. Food and Drug Administration (FDA), it will be the eighth pairing of the drug types OK'd in the past 4 years. More than 200 clinical trials are now testing this approach in various kinds of cancers, spurred by evidence that angiogenesis inhibitors help cancer-fighting T cells find their way deep into tumors. "There's a very clear rationale" for combining these two classes of drugs, says Robert Kerbel, a tumor immunologist at the University of Toronto.

This resurgence for angiogenesis inhibitors comes after astronomical expectations for the drugs were dashed in the late 1990s. Their premise was compelling. Tumors, hungry for oxygen and nutrients, would have to stimulate new blood vessels to sprout and grow to them; by preventing these vessels from forming, angiogenesis inhibitors should starve tumors and curb



Drugs that prevent new blood vessel growth may make tumors more vulnerable to checkpoint inhibitors.

PHOTO: ELLA MARU STUDIO/SCIENCE SOURCE

their growth. But the drugs fell flat for various reasons, including tumors becoming resistant or using other mechanisms to obtain blood. "It was an almighty crash" when the drugs didn't deliver, says University of Oxford tumor pathologist Francesco Pezzella, whose team discovered that tumors can commandeer existing vessels instead of growing new ones.

Still, angiogenesis inhibitors have recorded some successes. The monoclonal antibody bevacizumab gained regulatory approval in 2004 for patients with advanced colon cancer, and more than a dozen other angiogenesis inhibitors have since joined the antitumor arsenal. But on their own, angiogenesis inhibitors typically add only a few months to patients' lives. They "are very good at controlling disease, but by themselves they are not curative," says medical oncologist Brian Rini of Vanderbilt University Medical Center.

Scientists pair approved cancer drugs all the time to test whether they work better in concert, but a phenomenon discovered more than 25 years ago may explain why the checkpoint inhibitor-angiogenesis inhibitor partnership is productive. The blood vessels inside a tumor are a mess—bloated, twisting, and leaky. In the late 1990s, Rakesh Jain, a tumor biologist at Harvard Medical School (HMS), noticed that angiogenesis inhibitors had a surprising effect on tumors. The drugs "normalized" the vessels, spurring them to narrow, straighten, and become less porous. "What I saw was that the blood vessels [inside tumors] were getting better," says Jain, who published his "vascular normalization" hypothesis in 2001. Although controversial at first, the explanation is now widely accepted, says Oxford cancer biologist Anette Magnussen.

Abnormal blood vessels leave the interior of a tumor low on oxygen, inhibiting any T cells that enter to attack the tumor cells. That handicaps checkpoint inhibitors, which work by keeping cancer cells from flipping the inhibitory switches on T cells. By restoring normal circulation in the tumor, angiogenesis inhibitors may reverse the immune-suppressing conditions and permit tumor-targeting T cells to pounce. Jain says normalizing tumor vessels may provide another benefit—reining in metastasis, the spread of tumor cells to other parts of the body. An oxygen-starved tumor "is like a wounded tiger," he says. It is more dangerous in that state because it is prone to release cells that can relocate and establish new tumors elsewhere.

So far, dozens of clinical trials have evaluated checkpoint inhibitor-angiogenesis inhibitor combos. Some proved toxic, and others failed. However, the studies also led FDA to approve new therapies for types of liver, kidney, lung, and endometrial cancers. The drug combinations are not curative, but they restrain tumor growth. And some extend patients' lives by many months over angiogenesis inhibitors alone.

Trials have revealed other promising results as well, including those for HCC, which liver surgeon Pierce Chow of Duke-NUS Medical School in Singapore presented at the AACR meeting. In a phase 3 trial, he revealed, a combination of bevacizumab and the checkpoint inhibitor atezolizumab cut the odds that tumors would return after surgery or heat treatment by 28% in patients whose disease was diagnosed early. Researchers are still following the patients to determine whether these benefits persist and the combination increases survival.

> **"There's a very clear rationale" for pairing angiogenesis and checkpoint inhibitors.**
> **Robert Kerbel**, University of Toronto

Scientists are not just mixing and matching existing options. At least one startup, DynamiCure, is looking to develop new angiogenesis-thwarting drugs that are safer and can be coupled with checkpoint inhibitors. The company has already started a clinical trial of one antibody that stimulates vessel normalization. "When new companies are founded, that tells you the field is not dead, it's blossoming," Jain says.

Still, the strategy of combining the two drug types faces many challenges. For one thing, most angiogenesis inhibitors directly or indirectly stymie vascular endothelial growth factor (VEGF), a molecule that incites blood vessel growth. However, normal tissues also require VEGF, and the drugs can cause side effects such as bleeding, high blood pressure, and strokes.

Beyond side effects, researchers need better ways to determine whether tumors are responding to the drug duos, Magnussen says. "The challenge for using this [strategy] in humans is how you monitor blood vessel normalization in a cost-effective and time effective way."

HMS cancer biologist Dan Duda worries that researchers and pharmaceutical companies are rushing drug combinations into trials without considering the type of cancer and variables such as timing and schedule of treatments. "We are making the same mistakes we made the first time" with angiogenesis inhibitors, he says, adding, "We should take our time and think a little bit." ■

SCIENTIFIC COMMUNITY

Authorship listing for a harasser roils astronomy

After outcry, Geoff Marcy was removed from a recent manuscript. Will policy changes be next?

By **Katie Langin**

Within 2 weeks after a preprint was posted to arXiv on 31 March, outrage erupted on Twitter and multiple co-authors requested their names be removed. The issue wasn't the science, which focused on the detection of exoplanets, but its authorship. To the dismay of many astronomers, Geoff Marcy was listed as the third of 16 authors.

Marcy was once a prominent astronomer at the University of California (UC), Berkeley, known for leading efforts to search for planets in other solar systems, and he was key to the formation of a multi-investigator project called the Kepler Giant Planet Search (KGPS) that gathered the data behind the recent preprint. In 2015, *BuzzFeed News* published a bombshell story reporting details of a university investigation that found Marcy repeatedly violated sexual harassment policies. He resigned from his position later that year.

Marcy has published dozens of papers since then, but this one was different: After the outcry, he removed himself from the author list and agreed to not co-author any papers by the group going forward. That hasn't ended the matter, however. Many astronomers say the incident has galvanized discussion about what to do in similar situations in the future. "It would be extremely helpful if our professional societies would … have policies [stating] that people who had violated our code of conduct should not be allowed to be authors on papers," says Courtney Dressing, an astronomer at UC Berkeley. "That would take the decision burden out of individual research groups."

The recent paper drew from 13 years of KGPS data, some gathered by Marcy. "There's

their growth. But the drugs fell flat for various reasons, including tumors becoming resistant or using other mechanisms to obtain blood. "It was an almighty crash" when the drugs didn't deliver, says University of Oxford tumor pathologist Francesco Pezzella, whose team discovered that tumors can commandeer existing vessels instead of growing new ones.

Still, angiogenesis inhibitors have recorded some successes. The monoclonal antibody bevacizumab gained regulatory approval in 2004 for patients with advanced colon cancer, and more than a dozen other angiogenesis inhibitors have since joined the antitumor arsenal. But on their own, angiogenesis inhibitors typically add only a few months to patients' lives. They "are very good at controlling disease, but by themselves they are not curative," says medical oncologist Brian Rini of Vanderbilt University Medical Center.

Scientists pair approved cancer drugs all the time to test whether they work better in concert, but a phenomenon discovered more than 25 years ago may explain why the checkpoint inhibitor-angiogenesis inhibitor partnership is productive. The blood vessels inside a tumor are a mess—bloated, twisting, and leaky. In the late 1990s, Rakesh Jain, a tumor biologist at Harvard Medical School (HMS), noticed that angiogenesis inhibitors had a surprising effect on tumors. The drugs "normalized" the vessels, spurring them to narrow, straighten, and become less porous. "What I saw was that the blood vessels [inside tumors] were getting better," says Jain, who published his "vascular normalization" hypothesis in 2001. Although controversial at first, the explanation is now widely accepted, says Oxford cancer biologist Anette Magnussen.

Abnormal blood vessels leave the interior of a tumor low on oxygen, inhibiting any T cells that enter to attack the tumor cells. That handicaps checkpoint inhibitors, which work by keeping cancer cells from flipping the inhibitory switches on T cells. By restoring normal circulation in the tumor, angiogenesis inhibitors may reverse the immune-suppressing conditions and permit tumor-targeting T cells to pounce. Jain says normalizing tumor vessels may provide another benefit—reining in metastasis, the spread of tumor cells to other parts of the body. An oxygen-starved tumor "is like a wounded tiger," he says. It is more dangerous in that state because it is prone to release cells that can relocate and establish new tumors elsewhere.

So far, dozens of clinical trials have evaluated checkpoint inhibitor-angiogenesis inhibitor combos. Some proved toxic, and others failed. However, the studies also led FDA to approve new therapies for types of liver, kidney, lung, and endometrial cancers. The drug combinations are not curative, but they restrain tumor growth. And some extend patients' lives by many months over angiogenesis inhibitors alone.

Trials have revealed other promising results as well, including those for HCC, which liver surgeon Pierce Chow of Duke-NUS Medical School in Singapore presented at the AACR meeting. In a phase 3 trial, he revealed, a combination of bevacizumab and the checkpoint inhibitor atezolizumab cut the odds that tumors would return after surgery or heat treatment by 28% in patients whose disease was diagnosed early. Researchers are still following the patients to determine whether these benefits persist and the combination increases survival.

> **"There's a very clear rationale" for pairing angiogenesis and checkpoint inhibitors.**
>
> **Robert Kerbel**, University of Toronto

Scientists are not just mixing and matching existing options. At least one startup, DynamiCure, is looking to develop new angiogenesis-thwarting drugs that are safer and can be coupled with checkpoint inhibitors. The company has already started a clinical trial of one antibody that stimulates vessel normalization. "When new companies are founded, that tells you the field is not dead, it's blossoming," Jain says.

Still, the strategy of combining the two drug types faces many challenges. For one thing, most angiogenesis inhibitors directly or indirectly stymie vascular endothelial growth factor (VEGF), a molecule that incites blood vessel growth. However, normal tissues also require VEGF, and the drugs can cause side effects such as bleeding, high blood pressure, and strokes.

Beyond side effects, researchers need better ways to determine whether tumors are responding to the drug duos, Magnussen says. "The challenge for using this [strategy] in humans is how you monitor blood vessel normalization in a cost-effective and time effective way."

HMS cancer biologist Dan Duda worries that researchers and pharmaceutical companies are rushing drug combinations into trials without considering the type of cancer and variables such as timing and schedule of treatments. "We are making the same mistakes we made the first time" with angiogenesis inhibitors, he says, adding, "We should take our time and think a little bit." ■

SCIENTIFIC COMMUNITY

Authorship listing for a harasser roils astronomy

After outcry, Geoff Marcy was removed from a recent manuscript. Will policy changes be next?

By **Katie Langin**

Within 2 weeks after a preprint was posted to arXiv on 31 March, outrage erupted on Twitter and multiple co-authors requested their names be removed. The issue wasn't the science, which focused on the detection of exoplanets, but its authorship. To the dismay of many astronomers, Geoff Marcy was listed as the third of 16 authors.

Marcy was once a prominent astronomer at the University of California (UC), Berkeley, known for leading efforts to search for planets in other solar systems, and he was key to the formation of a multi-investigator project called the Kepler Giant Planet Search (KGPS) that gathered the data behind the recent preprint. In 2015, *BuzzFeed News* published a bombshell story reporting details of a university investigation that found Marcy repeatedly violated sexual harassment policies. He resigned from his position later that year.

Marcy has published dozens of papers since then, but this one was different: After the outcry, he removed himself from the author list and agreed to not co-author any papers by the group going forward. That hasn't ended the matter, however. Many astronomers say the incident has galvanized discussion about what to do in similar situations in the future. "It would be extremely helpful if our professional societies would … have policies [stating] that people who had violated our code of conduct should not be allowed to be authors on papers," says Courtney Dressing, an astronomer at UC Berkeley. "That would take the decision burden out of individual research groups."

The recent paper drew from 13 years of KGPS data, some gathered by Marcy. "There's



harasser. "Having my name associated with his professionally is bad for my career," says co-author Cayla Dedrick, a Ph.D. student studying astronomy at Pennsylvania State University. Marcy's inclusion also led to potential psychological harm, Dedrick says. "A lot of people in astronomy, especially a lot of women, are survivors of sexual assault and sexual harassment themselves, so seeing your name next to his—seeing his name at all—can be extremely triggering for a lot of people."

Many had hoped the 2021 incident would prevent Marcy's name from appearing on future papers. "I don't know how we're in this position again," a professor involved in the KGPS project told *Science* on the condition of anonymity. But they're hopeful that the latest controversy might spark change.

The editor-in-chief of AAS Journals, Ethan Vishniac, says he has limited power to ban authors from publishing in the journals he oversees. "We would only ban an author for reasons that had nothing to do with specific publication issues if we were told to do so by the society," he says. But the latest incident has galvanized discussions within AAS about what to do in these kinds of situations going forward, he adds. In an email to *Science*, AAS President Kelsey Johnson confirmed the society's ethics working group is now considering whether to classify "sexual harassment—and indeed all forms of harassment, discrimination, and bullying—as 'research misconduct.'" Such a move, she noted, "would strengthen the ability … to issue sanctions."

Researchers *Science* spoke with generally support such a change, though some note those discussions are likely to confront situations less clear-cut than Marcy's. It will be challenging to come up with an absolute rule as the severity of cases will vary, they say, and the society will have to decide what to do when the results of university investigations aren't made public.

In the absence of society action, researchers can take matters into their own hands, says Bruce Macintosh, an astronomer at UC Santa Cruz and the director of the UC Observatories. In his view, any action an individual takes to undermine a project—including harassing students and driving them out of the field—should be weighed against their positive contributions when making authorship decisions. Collaborative groups, he adds, can make their policies explicit by developing authorship guidelines that clearly state what happens if a team member commits such violations. "It's absolutely necessary that these large collaborations have rules," he says.

When Marcy's misconduct came to light 8 years ago, he says, "I think it was a start of a wake-up moment for astronomy. But we have a long way to go." ■

no question in my mind that Geoff did contribute to the early days of the project," one co-author told *Science* on the condition of anonymity.

But some argue that reports of Marcy's behavior, which include allegations that he kissed and groped students without their consent, should have trumped whatever scientific contributions he made. "He was actually found guilty according to a Title IX investigation. That almost never happens in academia," says Mia de los Reyes, a postdoc studying astrophysics at Stanford University and one of the first people on Twitter to question Marcy's inclusion on the author list. "If the astronomy community cannot do the right thing … in this extremely clear-cut case, then I'm beyond disappointed."

Graduate students who contributed to the KGPS project but were not included as authors expressed dismay that the paper made use of data they had collected, which effectively meant their "hard work and late nights" had bolstered a "serial sexual harasser's h-index." "I spent 18.3 hours of my life staying up late collecting data to unknowingly help prop him up," one student tweeted.

The outcry—which included criticism directed at Marcy's co-authors—spurred a reckoning for some of the researchers whose names appeared on the preprint. Before publishing, one had viewed Marcy's authorship as a matter of historical record, taking into account his contribution. But after it was posted online, the author—who spoke to *Science* on the condition of anonymity—learned others saw it differently. "It created the appearance … of people being accepting and tolerating behavior that clearly should not be accepted or tolerated," the author says.

As a result, that author and at least three others contacted first author and KGPS principal investigator (PI) Lauren Weiss, an astronomer at the University of Notre Dame and a former Ph.D. student of Marcy's. They requested their names be removed from the manuscript the group had submitted to AAS Journals, published by the American Astronomical Society (AAS). After contacting the handling editor, Weiss decided to withdraw the manuscript until the author list could be sorted out. She also posted a notice on arXiv: "It has come to my attention that there are significant concerns about the author list. … Accordingly, I am revisiting [it]."

In an email to *Science*, Weiss confirmed that Marcy's name won't appear on the manuscript when it's resubmitted. "His substantial early contributions to the project will be clarified in the acknowledgments. I also am revisiting the author list with the intention of extending authorship invitations more extensively," she wrote. One of the project's other lead PIs—Andrew Howard, an astronomer at the California Institute of Technology and a former postdoc of Marcy's—also told *Science* that Marcy won't co-author any papers the group publishes going forward. "The … team does not condone sexual harassment or abusive behavior toward any individual," he wrote. (Marcy did not answer specific questions but wrote via email: "I retired 8 years ago, and have been living peacefully with my wife of 29 years … dabbling in projects.")

The move to drop Marcy as a co-author is welcome news to some—but they are frustrated it took so long. A similar controversy arose in 2021, when the same research group published a series of papers with Marcy as an author. In that case, some of the co-authors were early-career scientists and Marcy's inclusion effectively meant they had to choose between getting credit for their work and signing their name next to a known sexual



Virgo spreads its 3-kilometer arms near Pisa, Ital

ASTROPHYSICS

Damage sidelines Virgo gravitational wave detector

The hunt for sources resumes this month with just the twin LIGO detectors and a limited science reach

By **Adrian Cho**

Later this month, physicists will resume their hunt for astrophysical monsters: black holes and neutron stars going bump in the dark and emitting ripples in space called gravitational waves. But one of the three detectors that have spotted such waves—Virgo, near Pisa, Italy—has run into technical problems that will delay its restart, 3 years after all the facilities shut down for maintenance and upgrades. For the next few months, just the two detectors of the Laser Interferometer Gravitational-Wave Observatory (LIGO), in Louisiana and Washington state, will take data, making it harder to pinpoint sources on the sky.

The problem appears to originate not in the upgrades, but in older parts that are creating noise that would drown out many signals, says Fiodor Sorrentino, a physicist with Italy's National Institute for Nuclear Physics (INFN) and Virgo's commissioning coordinator. "But we cannot be 100% sure" before opening the detector, he says. Daniel Holz, an astrophysicist at the University of Chicago, says such hiccups are normal, although LIGO and Virgo had dodged them. "We're owed this kind of bad luck because our excessive good luck had to run out."

The good luck began in 2015, when the LIGO detectors first sensed ripples produced when two massive black holes swirled into each other and merged. Two years later, LIGO and Virgo spotted a nearby merger of two neutron stars, which set off an explosion called a kilonova that was viewed by myriad telescopes as well. So far, the three detectors have tallied more than 90 mergers of black holes and two of neutron stars.

Each detector is a huge L-shaped optical device called an interferometer. Light bounces between weighty mirrors at the ends of each arm of the L. Some light leaks through the mirrors at the elbow, and the two light beams interfere, either canceling or reinforcing each other, depending on the arms' relative lengths. A passing gravitational wave generally stretches one arm more than the other, causing light to warble out of the device in sync with the wave.

To spot the minuscule stretching, the arms must be long. LIGO's extend 4 kilometers and Virgo's, 3 kilometers. The detectors must also squelch other vibrations to stabilize the length of each arm to 1 femtometer, the width of a proton. So the whole rig resides in a vacuum chamber, and an elaborate suspension system supports each mirror. Virgo's problems seem to have arisen in the suspension and the mirrors.

Each of its 40-kilogram mirrors hangs from a pair of thin glass fibers. In November 2022, a fiber supporting one mirror broke. Although the mirror fell a minimal distance, the jolt appears to have loosened one of four magnets affixed to the mirror and used to stabilize it, Sorrentino says. The magnet's movements generate a smidgen of heat—literally vibrations in the glass. Also, a mirror in the other arm that suffered a similar fall in 2017 now appears to have small internal crack that is growing and generating heat. The noise limits Virgo's sensitivity to roughly half what it was at the end of the last run.

The problems became apparent only recently because some of the upgrades took longer to commission than expected, says Gianluca Gemme, a physicist at INFN and spokesperson for the 850-member Virgo team. Rather than restart the detector, researchers will open its vacuum chamber to remove the loose magnet from one mirror and replace the other mirror. That work should be completed by July, Gemme says. Tuning the instrument would take a few months more. "If everything goes well and there are no additional hidden sources of noise, we should be able to join [LIGO] in the fall," Gemme says. Still, Sorrentino cautions, "This current situation is a bit scary because you never know what will happen when you put hands on your [mirrors]."

The two LIGO detectors are performing well and should be ready for the 24 May restart, says Patrick Brady, an astrophysicist at the University of Wisconsin-Milwaukee and spokesperson for the LIGO collaboration. But the temporary loss of Virgo will limit the science that can be done. Three detectors can pinpoint a source in the sky to within a few dozen square degrees. With two, the localization is far worse.

Virgo proved the value of such triangulation in August 2017 when it and the LIGO detectors spotted the first neutron star merger. Coordinates were quickly dispatched to astronomers, enabling radio dishes, optical telescopes, gamma ray detectors, and other instruments to home in on the explosion and detect the heavy elements it forged—making it literally a cosmic gold mine.

But even with just LIGO, the 18-month run should yield plenty of science, Brady says. LIGO's detectors are already 30% more sensitive than before and should spot a black hole merger once every 2 or 3 days, bagging about 270 in total. That haul should nail down the distribution of black hole masses and may reveal unusual mergers, such as between black holes spinning in different directions. That information could help reveal how black hole pairs form, Brady says. "The excitement is going to come in more than just increasing the number, but in getting these exceptional events." ■

BIOMEDICINE

Solvent exposure strongly linked to Parkinson's

Military recruits exposed to TCE decades ago have increased risk of the brain disease

By **Meredith Wadman**

A groundbreaking study has produced the most compelling epidemiological evidence yet that exposure to the chemical solvent trichloroethylene (TCE)—common in soil and groundwater—increases the risk of developing Parkinson's disease. The movement disorder afflicts about 1 million people in the United States and is likely the fastest growing neurodegenerative disease in the world; its global prevalence has doubled in the past 25 years.

The authors of the report, published this week in *JAMA Neurology*, examined the medical records of tens of thousands of Marine Corps and Navy personnel who trained at Marine Corps Base Camp Lejeune in North Carolina from 1975 to 1985, when water there was heavily contaminated with TCE. Camp Lejeune veterans had a 70% higher risk of developing Parkinson's disease decades later compared with similar veterans who trained elsewhere. The North Carolina contingent also had higher rates of symptoms such as erectile dysfunction and loss of smell that are early harbingers of Parkinson's, which causes tremors; problems with moving, speaking, and balance; and in many cases dementia. Swallowing difficulties often lead to death from pneumonia.

Only about 10% of Parkinson's cases can be explained by genetics, and environmental chemicals, including several pesticides and TCE, have been implicated as triggers for some of the remainder. The new study, led by researchers at the University of California, San Francisco (UCSF), represents by far the strongest environmental link between the solvent and the disease. Until now, the entire epidemiological literature included fewer than 20 people who developed Parkinson's after known TCE exposure.

The Camp Lejeune analysis "is exceptionally important," says Briana De Miranda, a neurotoxicologist at the University of Alabama at Birmingham. "It gives us an extremely large population to assess a risk factor in a very carefully designed epidemiological study."

"We had suspicions, but this is the evidence," agrees Gary Miller, a neurotoxicologist at Columbia University. "It's very compelling."

TCE is a colorless liquid that readily crosses biological membranes. Animal studies have shown that TCE inhibits a key energy-generating enzyme in an area of the midbrain responsible for movement control, destroying dopamine-producing neurons there, the same region where such cells die in the human disease. The solvent turns into vapor quickly and can be unknowingly inhaled, absorbed through skin or ingested.

In the 20th century TCE was widely used including for decaffeinating coffee, dry cleaning, carpet cleaning and, when inhaled, as an obstetrical anesthetic. Some of TCE's uses have been phased out, but it continues to be used in producing refrigerants and to clean and degrease heavy machinery, including by the military. It persists in soil and groundwater, and inhalation of vapor from these hidden sources is likely the prime route of exposure today.

Decades ago at Camp Lejeune, wells on the base were contaminated from leaking underground storage tanks, industrial spills, and waste disposal sites. Water used on the base contained TCE levels more than 70 times the level allowed by the U.S. Environmental Protection Agency (EPA). Recruits could have ingested TCE in food or water, been exposed through their skin when bathing or showering, or inhaled the highly volatile compound.

The researchers calculated the rate of Parkinson's disease in the Camp Lejeune veterans and compared it with the rate in more than 72,000 veterans who trained during the same period at Marine Corps Base Camp Pendleton in California where levels of TCE were lower. Analyzing the health records of nearly 85,000 Marine Corps and Navy personnel stationed at the North Carolina base for at least 3 months during the 10-year period starting in 1975, the UCSF team found that by 2021, 279 of the Camp Lejeune veterans, or 0.33%, had developed Parkinson's versus 151, or 0.21%, of the control group. After adjusting for differences in age, sex, race, and ethnicity, the scientists found veterans from Camp Lejeune had a 70% higher rate of Parkinson's disease than the Camp Pendleton group.



Exposure to contaminated water at a military training base seems to have increased the risk that recruits would later develop Parkinson's disease.

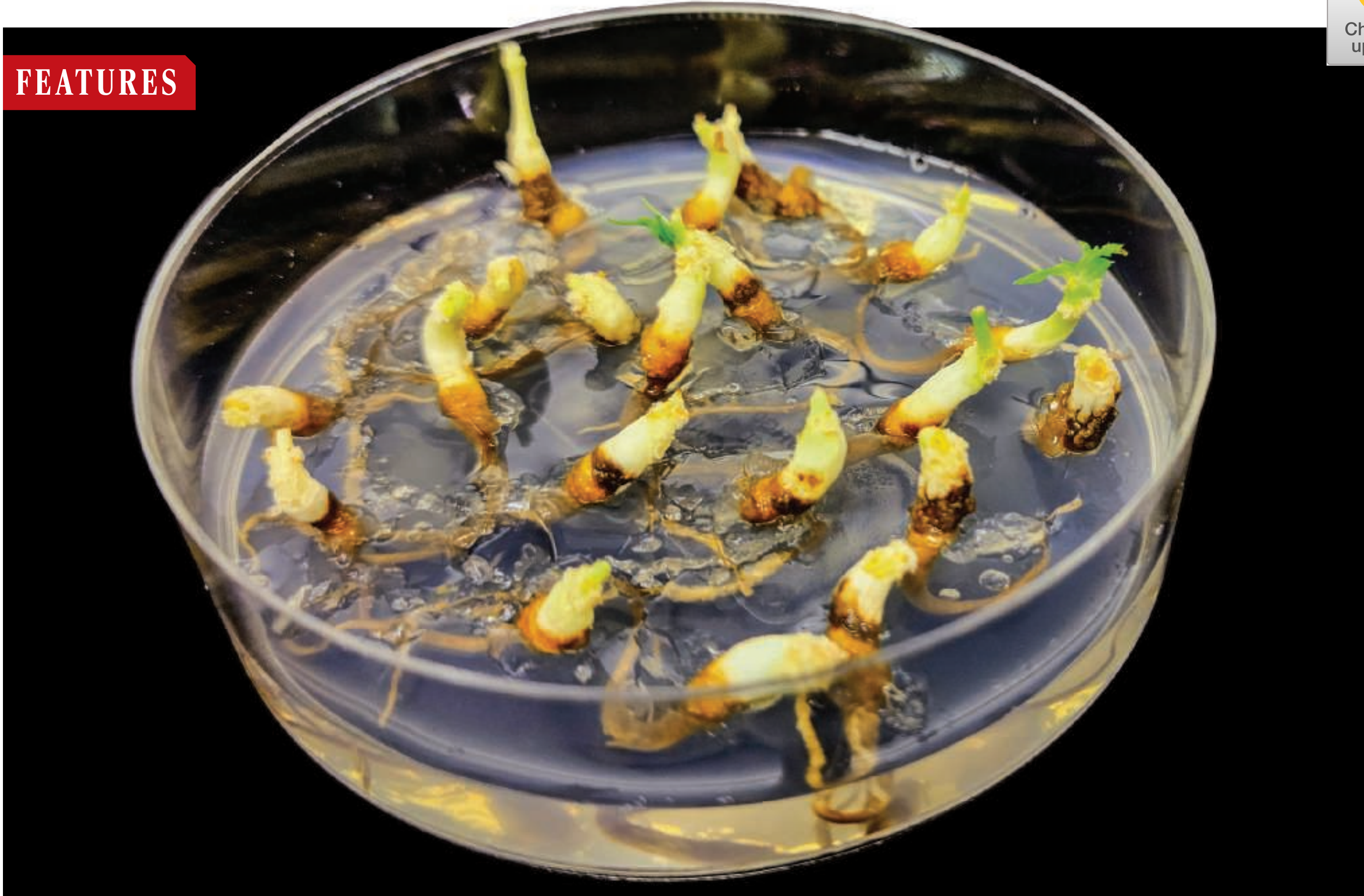
PHOTO: ALLEN BREED/AP

The new study's lead author, UCSF epidemiologist Sam Goldman, conducted a small twin study published in 2012 showing that TCE exposure increased the risk of the disease. He was motivated to undertake the current study in 2017 when the U.S. government declared that anyone who served at Camp Lejeune in the contaminated water era and developed Parkinson's disease would be entitled to benefits from the Department of Veterans Affairs on the assumption TCE had caused the condition. "I just felt … we really need to have greater certainty about this," Goldman says.

In January, EPA declared that TCE presents an "unreasonable risk of injury to human health" and said it will develop a rule regulating its use. (The chemical is also a known carcinogen.) But that "really means nothing for what's already in the environment," De Miranda says. Mitigating against exposure is tricky, she adds, because underground TCE locations aren't always documented.

The Camp Lejeune study will likely add ammunition to class action lawsuits that were launched after Congress last year enabled veterans from the base to sue the government for health damage they suffered from exposure to the contaminated water there decades ago. "This is increasing evidence that environmental factors are important causes of Parkinson's disease," Miller says. "But we are just scratching the surface. We need to continue studying this." ■

FEATURES



Scientists hope to alter cowpea seedlings so the crop can reproduce clonally.

UNMIXED BLESSING

Researchers are closing in on the long-sought goal of creating high-yielding hybrid crops that can be propagated indefinitely, without sex

By **Erik Stokstad**

In early summer, unusual pollinators swoop over rice fields in Texas and Arkansas. Small, nimble helicopters fly low and steady so their rotors blow pollen from one row of plants to another. The flights help RiceTec, a plant breeding company, produce seed for high-yielding, robust varieties of rice grown across the southern United States. It's an expensive and complicated way to create seed.

But the effort is worthwhile because the seeds sprout into plants with a mysterious robustness and resilience. The phenomenon, called hybrid vigor, comes from crossing two strains of inbred parents. Why hybrids are superior to normal plants is not clear, but one long-standing hypothesis is that favorable versions of genes from one parent dominate poor-performing, recessive genes from the other.

The development of hybrid varieties has boosted the yield of maize, sorghum, and other crops by up to 50% and has resulted in other valuable traits, such as better drought tolerance. But the method is only feasible in some species; there's no practical way to produce hybrid wheat or soybeans, for example. And when it works, it's extremely labor intensive.

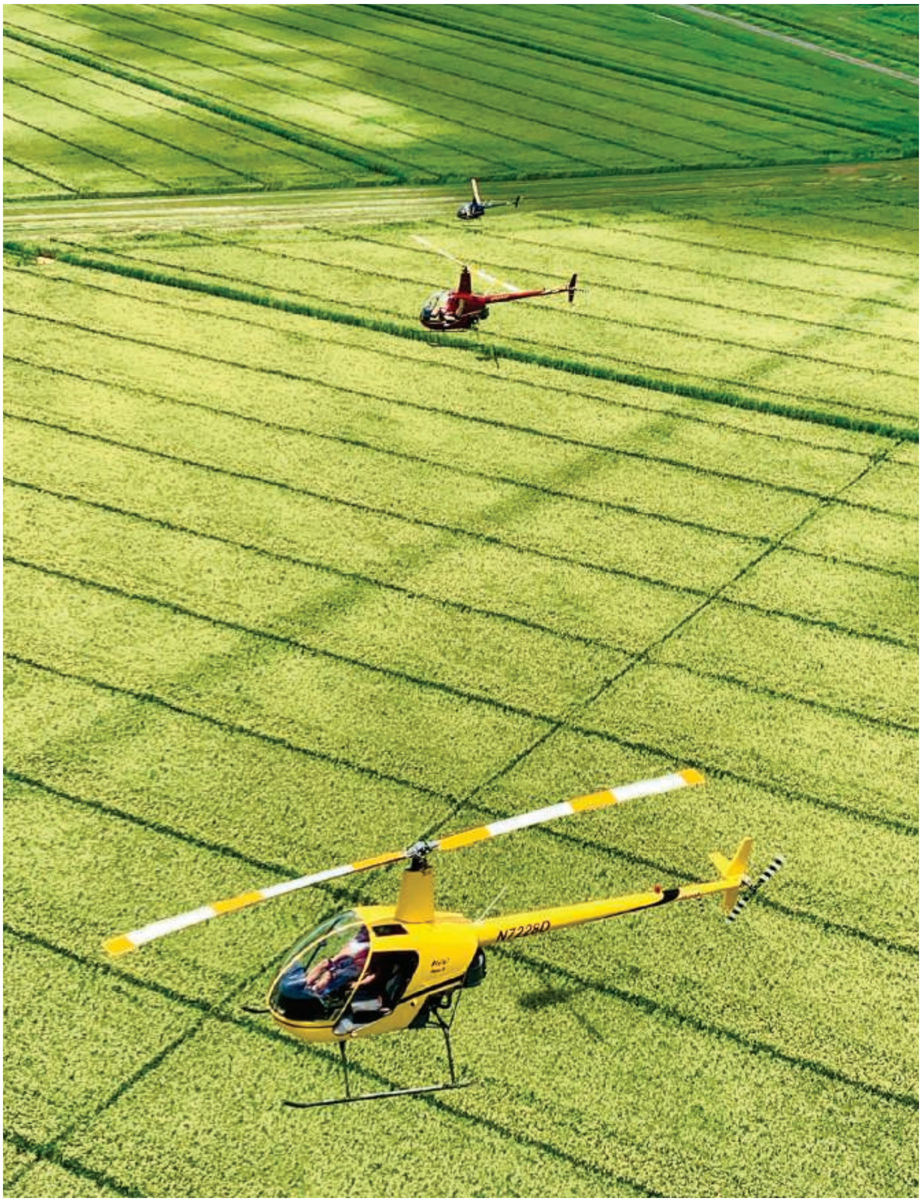
In rice, seed companies must first develop a strain of plants that can't self-pollinate. Then come the helicopters, which sweep in pollen from a second strain. The process has to be repeated for each new batch of seed to avoid the reshuffling of genes and loss of favorable traits that happens during ordinary sexual reproduction. "It's a very imperfect system," says José Ré, vice president of research at RiceTec.

Plant breeders have long dreamed of an easier, more powerful way to create hybrid seed. In nature, some plant species reproduce clonally: The eggs inside their flowers become embryos without pollination, part of a process called apomixis—"away from mixing" in Greek. If researchers could genetically engineer crops to reproduce through apomixis, the process of creating the first hybrid generation might still be laborious. But then seed companies could much more easily propagate hybrid offspring.

For decades, scientists had limited success. But recent breakthroughs have brought

the concept closer to reality. In 2019, an international team reported that it had successfully engineered a line of rice plants that could reproduce clonally—the first instance of synthetic apomixis in a crop. Groups around the world are working to develop apomictic varieties of sorghum, tomatoes, alfalfa, and other crops. There's a palpable "sense of excitement" in the field, says Mary Gehring, a molecular biologist at the Whitehead Institute and the Massachusetts Institute of Technology who studies development in apomictic plants.

The technology won't be ready to be commercialized for years. "There's still an awful lot that we don't understand about how to make it efficient for agriculture," says Peggy Ozias-Akins, a geneticist at the University of Georgia. But seed companies are paying attention. Apomictic reproduction would simplify how they produce hybrid seeds, quicken the release of new varieties, and reduce costs. The technology could also benefit smallholder farmers in poorer countries who might not have regular access to commercial hybrid seeds, because they could save seeds produced by the previous year's crop. "It really would be a big game changer," says Adam Famoso, a rice breeder at Louisiana State University.



Some methods of hybrid rice production employ helicopters to blow pollen from one rice strain to another.

PHOTO: RICK SÁ

THE DISCOVERY OF VIRGIN birth in plants is widely credited to John Smith, a 19th century botanist who served as the inaugural curator at London's Royal Botanic Gardens. For a decade, he had watched three holly plants from Australia bear fruit without having ever produced a male flower or anything that resembled pollen, the vehicle for plant sperm. In 1839, Smith reported to the Linnean Society of London that he could grow new plants from the seeds of the hollies. It was an incendiary claim that was met with "incredulity," according to Thomas Meehan, a botanist writing decades later.

In 1898, however, Swedish botanist Oscar Juel demonstrated, with convincing microscopy, that the egg cells of a plant called alpine catsfoot could develop into embryos in the absence of pollen. Other researchers took a closer look at their own favorite species. As evidence accumulated, more and more botanists began to take the phenomenon seriously. Today, apomictic reproduction has been confirmed in more than 400 plant species but no staple crops.

In the late 1990s, as genetic tools became more readily available, experts were optimistic they could identify the genes behind apomixis and deploy them in crops, creating clones that would bypass the genetic recombination that happens during plant sex, which shuffles away favorable gene combinations. But progress was slow. "People said, 'OK, we will crack that nut,'" recalls Erik Jongedijk, a plant geneticist at KWS, a major seed company in Europe. And for a long time, "it was never cracked."

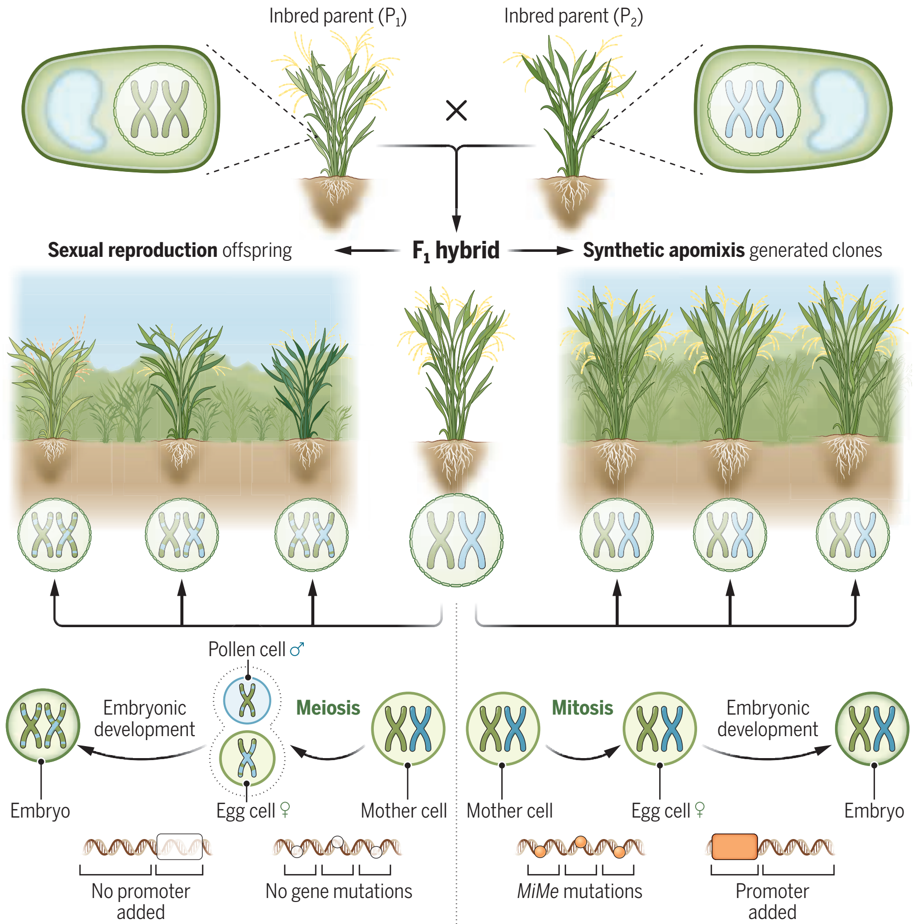
Part of the difficulty stems from the complexity of the reproductive process researchers are trying to modify. During sexual reproduction, gametes—eggs and sperm—are created through meiosis, a process that results in haploid cells, with half the number of chromosomes. To form embryos with a full complement of chromosomes, eggs and sperm need to come together. Many naturally apomictic plants instead create gametes through mitosis, with no change in the chromosome count. The eggs can then turn into an embryo without being fertilized, in a process known as parthenogenesis.

It's taken decades to identify some of the genes involved and to figure out how to tinker with them. In 2009, a group of scientists led by Raphaël Mercier, a geneticist now at the Max Planck Institute for Plant Breeding Research, showed that if they knocked out three genes involved in meiosis in the model plant *Arabidopsis*, it would make gametes through mitosis, preserving their full set of chromosomes. They named the trio of mutations *MiMe*, short for "mitosis instead of meiosis." In 2016, they replicated the feat in rice, showing that the *MiMe* mutations would create diploid eggs genetically identical to the mother plant.

Meanwhile, other groups were figuring out how to coax egg cells to develop into embryos without being fertilized. In 2002, geneticist Kim Boutilier—now at Wagenin-

Engineering perpetual vigor

Hybrid rice is prized by farmers because it is robust and high yielding. Breeders produce it by creating strains of highly inbred parents and crossing them. But only first generation (F_1) crosses have the prized genetics; sexual reproduction shuffles away the favorable traits and renders the seeds of later generations less valuable (below, left). To get around that problem, scientists have genetically engineered a rice strain that reproduces clonally, through a process called apomixis (below, right).



Sexual reproduction
In sexual plants, eggs and pollen are created through meiosis, a type of cell division that halves the number of chromosomes and creates new gene combinations through recombination. Sperm and egg combine to yield an embryo that has the full complement of chromosomes.

Synthetic apomixis
Adding *MiMe* mutations allows the plant to create egg cells via mitosis, giving them a full set of chromosomes and avoiding recombination. Inserting a promoter turns on a gene called *BABY BOOM* in the egg cells, triggering embryo development without fertilization.

gen University—made what turned out to be a key discovery when she and her colleagues identified a gene called *BABY BOOM* in rapeseed. The gene, the team found, triggered the growth of embryos from shoots and leaves when it was turned on in *Arabidopsis*.

Since then, *BABY BOOM*–like genes have been found in many plants. In 2015, Ozias-Akins and her colleagues identified the function of one of them in a natural apomict, the grass *Pennisetum squamulatum*. When they transferred this gene into a closely related grass that reproduces sexually, as well as into rice and maize, it induced parthenogenesis, resulting in haploid embryos that were viable. "Amazingly, it worked," says Ueli Grossniklaus of the University of Zurich, a developmental geneticist who studies reproduction in *Arabidopsis* and natural apomicts.

The stage was set for the next step of the engineering process: knocking out the trio of meiosis genes and activating the critical *BABY BOOM* gene within a single plant.

SCIENTISTS DIDN'T WANT to simply transfer the key embryo activating gene from a grass into rice containing *MiMe* mutations. Although such a step is technically possible, it comes with a regulatory downside: Crops that are created by transferring a gene between species require lengthy regulatory evaluation before they can go to market. So scientists looked to rice's own genome instead.

A landmark accomplishment came from someone who didn't set out to study apomixis. Venkatesan Sundaresan, a development biologist at the University of California (UC), Davis, had been investigating genes expressed in rice as egg and sperm cells fuse and develop into embryos. He and his colleagues noticed that the rice version of *BABY BOOM* is normally expressed in a plant's sperm cells, but not the eggs, and that it remains active in the embryo after fertilization. If they could activate *BABY BOOM* in the egg cells as well, it could make pollination unnecessary, they reasoned.

To test whether that would work, UC Davis team member Imtiyaz Khanday isolated rice's *BABY BOOM* and tacked on a promoter that turns the gene on specifically in the egg. The next step was to get the gene back into the plants. Khanday inserted this DNA package into the genome of a plant-infecting bacterium, a standard vehicle for genetically modifying plants. Added to a petri dish of rice cells, the modified bacterium inserted its own DNA—including the added gene—into the crop's DNA. The cells formed a tumor-like tissue known as callus that could then be coaxed to grow into a seedling.

The resulting plants, which were also modified using CRISPR gene editing to have the *MiMe* mutations, underwent parthenogenesis and produced clonal seeds. Those seeds germinated and the seedlings were genetically identical to the mother plant, the team reported in 2019 in *Nature*.

It was a groundbreaking achievement, but the plants weren't perfect: Only about 30% of their seeds were clonal. (The remainder had been fertilized by pollen produced by the genetically modified plants and were not viable.) So the team continued to tinker with its methods. In a follow-up paper published in *Nature Communications* in December 2022, the researchers reported that after using a different variety of rice—a commercial hybrid—and adding the *MiMe* mutations and the *BABY BOOM* promoter in the same step, they ended up with plants that yielded more than 95% clonal seeds. It's "absolutely fantastic work," Jongedijk says.

The group is now looking for funding to test its clonal rice in field trials, says coauthor Emmanuel Guiderdoni, a rice geneticist at the French Agricultural Research Centre for International Development. The researchers want to see how their clonal plants fare in harsher conditions than they experienced in the greenhouse.

RESEARCH ON OTHER CROPS is also picking up steam. Anna Koltunow of the University of Queensland is developing apomictic varieties of sorghum and cowpea, important crops for farmers in sub-Saharan Africa. In



Venkatesan Sundaresan and his colleagues engineered rice plants to reproduce clonally, which could simplify hybrid seed production.

October 2022, she and her team began field trials in Australia of hybrid sorghum that was genetically modified to be parthenogenic and produce haploid embryos; they plan to add *MiMe* mutations to strains in the future. At least 10 groups in China are also working on apomictic varieties of cabbage, tomatoes, alfalfa, and other vegetable and forage crops, says Kejian Wang, a geneticist at the China National Rice Research Institute who is developing an apomictic strain of hybrid rice.

One challenge for those researching dicots—the large group of flowering plants that includes beans and vegetable crops—has been that *BABY BOOM* doesn't seem to work in them when expressed in egg cells. "We've tried really hard, and nothing ever worked," Grossniklaus says. But now they have another option. In January, a group led by plant geneticist Peter van Dijk of KeyGene, a plant breeding company, reported in *Nature Genetics* the discovery of *PAR*, a gene in dandelion—a naturally apomictic dicot—that appears to have a similar function as *BABY BOOM*. When his team switched on *PAR* in lettuce, rudimentary embryos formed without fertilization. The search for that gene took more than 15 years. "It's a beautiful story, but it also demonstrates how much work and effort was needed," says Tim Sharbel, an evolutionary biologist at the University of Saskatchewan.

Jongedijk expects another 5 to 10 years of research may be needed before synthetic apomixis can be deployed commercially in any crop. For an apomictic hybrid variety to appeal to farmers, 100% of the seeds must be clonal because they won't want the less vigorous seeds that are produced through normal sexual reproduction. And as with all new crop varieties, scientists will need to conduct extensive field testing to determine how hybrid varieties respond to drought and other stressors.

Further tinkering with genes that control seed development could lead to even more progress. Like most plants, the apomictic rice now under development still needs pollen to fertilize its endosperm—the seed tissue that provides sustenance for the developing embryo. For those strains—as well as for commercial varieties that reproduce sexually—that step is vulnerable to climate change, because pollen can become less viable when it's exposed to high temperatures. Gehring's team recently received funding to try to engineer plants that develop endosperm without being fertilized, a feat that some naturally apomictic plants are capable of. If the team is successful, future apomictic crop varieties wouldn't rely on pollen at all—enabling them to produce bountiful seeds even during heat waves.



Dandelions are natural apomicts. Their yellow blooms consist of many tiny flowers, each with clonal eggs.

As research on apomixis proceeds, some breeders caution that even when perfected, such crops might not succeed in the market. Famoso notes that some countries, especially in Asia, that have resisted genetically modified foods may not want to import apomictic rice. But Jauhar Ali, head of the hybrid rice program at the International Rice Research Institute, is more optimistic. "Gene editing is slowly being accepted and many governments are understanding the importance of this tool for bringing benefits to agriculture," he says.

"There's a lot of promise for this technology and hopefully it can make it into a farmer's field someday," Famoso says. For now, though, the helicopters will keep flying. ■

PHOTOS: (TOP TO BOTTOM) UNIVERSITY OF CALIFORNIA, DAVIS; DAVID J. GREEN/ALAMY

INSIGHTS

PERSPECTIVES



ANTHROPOLOGY

The ancient history of kissing

Sources from Mesopotamia contextualize the emergence of kissing and its role in disease transmission

By **Troels Pank Arbøll**[1] and **Sophie Lund Rasmussen**[2,3,4]

Recent studies maintain that the first known record of human romantic-sexual kissing originates in a Bronze Age manuscript deriving from South Asia (India), tentatively dated to 1500 BCE (*1*). Yet, a substantial corpus of overlooked evidence challenges this premise because lip kissing was documented in ancient Mesopotamia and Egypt from at least 2500 BCE onward. Because this behavior did not emerge abruptly or in a specific society but appears to have been practiced in multiple ancient cultures over several millennia, the kiss cannot be regarded as a sudden biological trigger causing a spread of specific pathogens, as recently proposed (*2*). Further understanding of the history of kissing in human societies—and its secondary effect on disease transmission—can be gained from a case study of sources from ancient Mesopotamia (modern-day Iraq and Syria).

In research, two types of kissing are generally differentiated, namely the friendly-parental kiss and the romantic-sexual kiss. Whereas friendly-parental kissing appears to be ubiquitous among humans across time and geography, romantic-sexual kissing is not culturally universal, and it is dominant in stratified societies (*3*). Research has suggested that romantic-sexual kissing evolved for the purposes of evaluating aspects of a potential mate's suitability through chemical cues communicated in the saliva or breath, mediating feelings of attachment between pair-bonded individuals, and facilitating sexual arousal and thereby sexual relations (*3*). Kissing is also attested in other animal

A clay model from Mesopotamia, dated ~1800 BCE, shows a couple kissing. The original is kept at the British Museum.

species, such as mouth-to-mouth kissing with a romantic-sexual purpose in bonobos (*Pan paniscus*) and platonic kissing to manage social relationships in chimpanzees (*Pan troglodytes*) (*1*). These two species constitute the closest living relatives to humans, and their practices of kissing may hint at the presence and evolution of this behavior in human ancestors (*4*).

In a study investigating the transfer of the oral microbe *Methanobrevibacter oralis*, it has been hypothesized that Neanderthals could have engaged in lip kissing with modern humans more than 100,000 years ago (*5*, *6*). Still, the advent of romantic-sexual kissing remains uncertain, although two prehistoric sculptures from Ain Sakhri (BM 1958,1007.1) and Malta (T/p1014) might imply its existence before the invention of writing.

Humanity's earliest recorded kiss occurs in sources from the ancient Middle East. Kissing is attested in ancient Mesopotamian texts from 2500 BCE onward. Ancient Mesopotamia constituted the areas along the Euphrates and Tigris rivers, which today roughly cover Iraq and Syria. Writing was first invented simultaneously in southern Iraq and in Egypt around 3200 BCE. In Mesopotamia people wrote in cuneiform script on clay tablets, which primarily recorded the Sumerian and Akkadian languages from ~3200 BCE to 75 CE. In the earliest texts in the Sumerian language, kissing was described in relation to erotic acts, possibly as a postcoital activity, and the locus was the lips (*7*). In the Akkadian language, references to kissing can be subdivided into two distinct groups, the first designating friendly and familial affection, describing a display of submission or respect through the act of kissing the feet or the ground, and the second being an erotic action with the lips as the primary locus (*7*).

Considering the thousands of cuneiform texts that are available, there are relatively few instances where romantic-sexual kissing is described. Regardless, there are clear examples illustrating that kissing was considered an ordinary part of romantic intimacy in ancient times. The texts imply that kissing was something that married couples did (*8*), though the kiss was regarded as part of an unmarried person's sexual desire when in love (*8*). Two texts from ~1800 BCE are especially revealing. One describes how a married woman was almost led astray by a kiss from another man, and the other describes an unmarried woman swearing to avoid kissing and having sexual relations with a specific man (*9*). Apparently, society tried to regulate such activities between unwed people or adulterers. Furthermore, the sexual aspect of kissing was frowned upon in public, and kissing a person who was not meant to be sexually active, such as a priestess, was believed to deprive the kisser of the ability to speak (*9*). Still, it seems that tokens of friendship and familial affection, such as that between mother and child, included kissing (*7*). Kissing was also used in ritual contexts, where a person in need of divine restoration could kiss a person in a state of trance, an old woman, or a slave girl.

Reaching beyond its importance for social and sexual behavior, the act of kissing may have played a secondary and unintentional role throughout history in facilitating the transfer of orally transmitted microorganisms, potentially causing disease.

"Humanity's earliest recorded kiss occurs in sources from the ancient Middle East."

Infectious diseases have been around since the dawn of history, with a constant evolutionary arms race unfolding between pathogens and hosts. Recent advances in the technology for extracting ancient DNA have enabled the detection of a wide range of pathogen genomes, such as herpes simplex virus 1 (HSV-1) (*2*), Epstein-Barr virus (*10*), and human parvovirus B19 (*11*), in ancient human remains. These pathogens can infect humans through a range of different transmission routes, including saliva, making any act of kissing a potential means of spreading infection (*12*). The confirmation of microbial genomes deriving from human remains dating back thousands of years indicates that potentially kiss-transmissible organisms were present in historical and even prehistorical periods (*5*).

Recently, an interesting study presented ancient European human HSV-1 genomes deriving from dental material from human skeletons dating from 253 to 1700 CE (*2*). The authors discovered that there was a shift in dominant lineages of HSV-1 in the Bronze Age, and they suggested that this shift occurred because of an additional route of lateral transmission of HSV-1, potentially linked to the introduction of new cultural practices caused by migration, such as romantic-sexual kissing. Given this example, it is possible that symptoms indicating HSV-1 were included in the substantial corpus of medical manuscripts that describe symptoms of illness from ancient Mesopotamia. Although this material should be cautiously approached when using retrospective diagnosis, which interprets ancient descriptions of a specific illness by correlating the symptoms with those of a modern disease, it cannot be ignored that the ancient medical texts were influenced by a variety of cultural and religious concepts. It must therefore be emphasized that the preserved texts cannot be read at face value.

It has been proposed that the disease called *bu'šānu*, which was described in ancient medical texts, might have reflected HSV-1 infection (*10*) in addition to a number of other modern diseases, such as diphtheria. The *bu'šānu* disease was located primarily in or around the mouth and pharynx, and the name itself was derived from a verb meaning "to stink" (*13*). Various descriptions of the disease exist, and several include the symptom *bubu'tu* (*14*). The interpretation that *bu'šānu* could cover HSV-1 infection is mainly based on the observation that *bubu'tu* might be interpreted as "vesicle," although the translations "pustule" and "boil" remain dominant. Vesicles in or around the mouth are one of the dominant signs of HSV-1 infection.

Ideas of how humans contracted disease in ancient Mesopotamia differed from the modern understanding (*15*), and people in Mesopotamia did not attribute *bu'šānu*, or the spread of infectious diseases in general, to kissing. Regardless, certain cultural and religious factors instituted a partial social control on this practice. Such regulation of intimate relations could have had the unrecognized benefit of reducing the spread of pathogens. Still, it seems that experience with contagious illness may have brought about other pragmatic means to avoid infection through bodily fluids. In a letter from ~1775 BCE, a woman in a palace harem had fallen ill with an infectious disease that caused lesions (*9*). To avert infection, everyone was instructed to avoid drinking from her cup, sleeping in her bed, or sitting on her chair.

The sources from ancient Mesopotamia suggest that kissing in relation to sex, family, and friendship was an ordinary part of everyday life in central parts of the ancient Middle East from the late 3rd millennium BCE onward. Accordingly, kissing should not be regarded as a custom originating exclusively in any single region and spreading from there (*7*). Furthermore, the sources from Mesopotamia show that the romantic-sexual kiss was known far

[1]Department of Cross-Cultural and Regional Studies, University of Copenhagen, Copenhagen, Denmark. [2]Linacre College, University of Oxford, Oxford, UK. [3]Wildlife Conservation Research Unit (WildCRU), Department of Zoology, University of Oxford, Recanati-Kaplan Centre, Tubney, UK. [4]Department of Chemistry and Bioscience, Aalborg University, Aalborg, Denmark.
Email: troels.arboell@gmail.com

earlier, and in a wider geographic area, than the references from India dated to 1500 BCE (*1*, *2*), which stands in contrast to previous observations about the history of kissing (*2*).

Accounting for an ancient cultural diffusion of the romantic-sexual kiss is challenging, and it seems to have had numerous independent origins. Although some societies may not have practiced the romantic-sexual kiss, it must have been known in most ancient cultures, which could be indicative of a near universality. The written documentation to support this observation originates in stratified and complex societies, and the question remains whether the findings are only typical of behaviors found in complex state-level societies. Evidence indicates that kissing was a common practice in ancient times, potentially representing a constant influence on the spread of orally transmitted microbes, such as HSV-1. It therefore seems unlikely that kissing would have arisen as an immediate behavioral adaptation in other contemporary societies, which inadvertently accelerated disease transmission. Finally, the debate about kissing as a vector of disease transmission illustrates the benefits of an interdisciplinary approach to produce a holistic representation of historical disease transmission through social interactions. ■

REFERENCES AND NOTES

1. S. Kirshenbaum, *The Science of Kissing: What Our Lips Are Telling Us* (Grand Central Publishing, 2011).
2. M. Guellil *et al*., *Sci. Adv*. **8**, eabo4435 (2022).
3. W. R. Jankowiak, S. L. Volsche, J. R. Garcia, *Am. Anthropol*. **117**, 535 (2015).
4. G. E. King, in *Encyclopedia of Animal Cognition and Behavior*, J. Vonk, T. Shackelford, Eds. (Springer, 2020), pp. 1–12.
5. L. S. Weyrich *et al*., *Nature* **544**, 357 (2017).
6. E. Callaway, *Nature* **543**, 163 (2017).
7. J. Cooper, in *Reallexikon der Assyriologie und Vorderasiatischen Archäologie. Sechster Band. Klagegesang – Libanon*, D. Edzard, Ed. (de Gruyter, vol. 6, 1980), pp. 375–379.
8. N. Wasserman, *Akkadian Love Literature of the Third and Second Millennium BCE* (Harrassowitz, vol. 4, 2016).
9. M. Stol, *Women in the Ancient Near East* (de Gruyter, 2016).
10. P. W. Ewald, *Int. J. Paleopathol*. **21**, 178 (2018).
11. B. Mühlemann *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, 7557 (2018).
12. J. Limeres Posse, P. Diz Dios, C. Scully, in *Saliva Protection and Transmissible Diseases* (Academic Press, 2017), pp. 53–92.
13. B. Böck, *The Healing Goddess Gula: Towards an Understanding of Ancient Babylonian Medicine* (Brill, 2014).
14. N. P. Heeßel, *Babylonisch-assyrische Diagnostik* (Ugarit-Verlag, 2000).
15. T. P. Arbøll, *Medicine in Ancient Assur: A Microhistorical Study of the Neo-Assyrian Healer Kişir-Aššur* (Brill, 2021).

ACKNOWLEDGMENTS

The authors acknowledge financial support from the Carlsberg Foundation and the Edubba Foundation.

10.1126/science.adf0512

ENGINEERING

The disappearing boundary between organism and machine

Artificial skin mimics the sensory feedback of biological skin

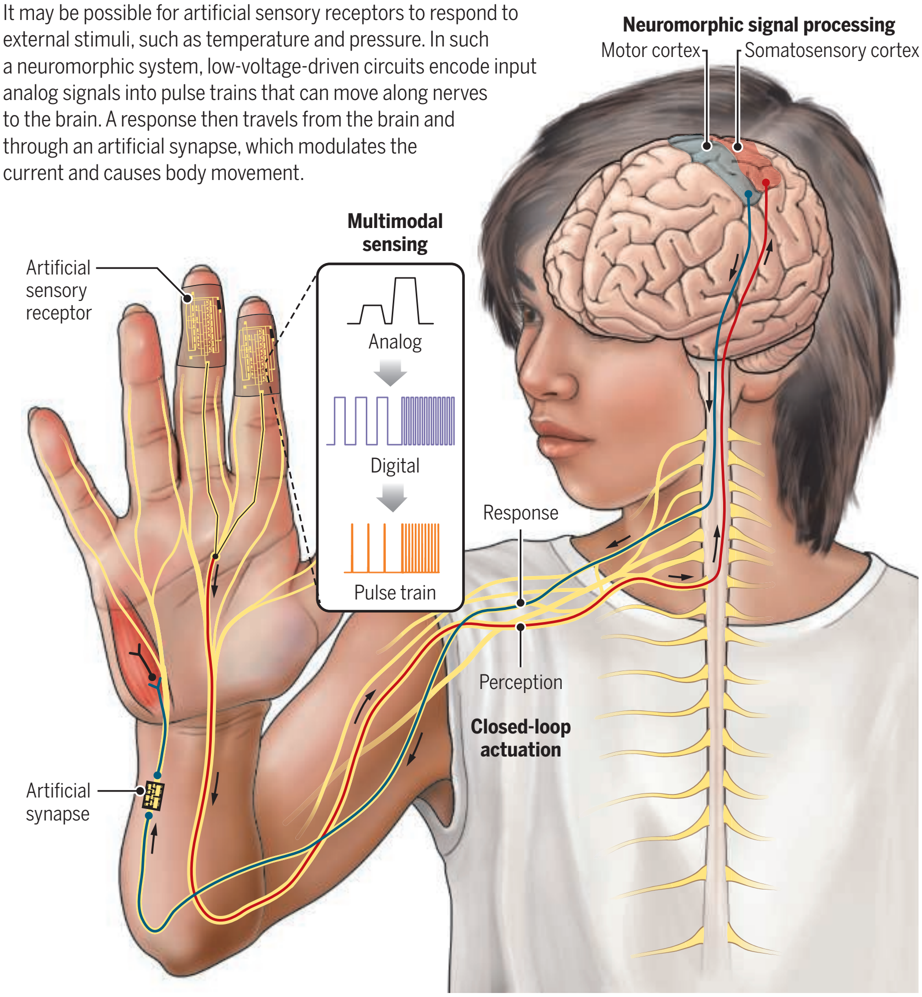
By **Tsuyoshi Sekitani**

Research and development in artificial intelligence (AI) and robotics continue to explore the interface of living organisms with machines. AI-based, motorized prosthetic arms and legs, as well as brain-machine interfaces that connect prosthetics to human brain activity, aim to eventually succeed in fully mimicking human bodily functions and serve as medical solutions for people suffering from illness and injury. However, to more seamlessly integrate mechanical actuators, such as prosthetics, with the human body, the development of a large, elaborate, and stretchable sensor that covers both the body and the mechanical components is necessary (*1*–*6*). Electronic skin (e-skin) would eliminate the boundary between the living body and machine components. On page 735 of this issue, Wang *et al.* (*7*) report the development of a neuromorphic e-skin system that mimics the sensory feedback of biological skin.

Mechanically, the body is a skin-covered sensor and an actuator that converts input energy into an intended physical motion. Sensory organs within the surface of the

An electronic skin that supports a sensorimotor loop

It may be possible for artificial sensory receptors to respond to external stimuli, such as temperature and pressure. In such a neuromorphic system, low-voltage-driven circuits encode input analog signals into pulse trains that can move along nerves to the brain. A response then travels from the brain and through an artificial synapse, which modulates the current and causes body movement.



earlier, and in a wider geographic area, than the references from India dated to 1500 BCE (*1*, *2*), which stands in contrast to previous observations about the history of kissing (*2*).

Accounting for an ancient cultural diffusion of the romantic-sexual kiss is challenging, and it seems to have had numerous independent origins. Although some societies may not have practiced the romantic-sexual kiss, it must have been known in most ancient cultures, which could be indicative of a near universality. The written documentation to support this observation originates in stratified and complex societies, and the question remains whether the findings are only typical of behaviors found in complex state-level societies. Evidence indicates that kissing was a common practice in ancient times, potentially representing a constant influence on the spread of orally transmitted microbes, such as HSV-1. It therefore seems unlikely that kissing would have arisen as an immediate behavioral adaptation in other contemporary societies, which inadvertently accelerated disease transmission. Finally, the debate about kissing as a vector of disease transmission illustrates the benefits of an interdisciplinary approach to produce a holistic representation of historical disease transmission through social interactions. ■

REFERENCES AND NOTES

1. S. Kirshenbaum, *The Science of Kissing: What Our Lips Are Telling Us* (Grand Central Publishing, 2011).
2. M. Guellil *et al*., *Sci. Adv*. **8**, eabo4435 (2022).
3. W. R. Jankowiak, S. L. Volsche, J. R. Garcia, *Am. Anthropol*. **117**, 535 (2015).
4. G. E. King, in *Encyclopedia of Animal Cognition and Behavior*, J. Vonk, T. Shackelford, Eds. (Springer, 2020), pp. 1–12.
5. L. S. Weyrich *et al*., *Nature* **544**, 357 (2017).
6. E. Callaway, *Nature* **543**, 163 (2017).
7. J. Cooper, in *Reallexikon der Assyriologie und Vorderasiatischen Archäologie. Sechster Band. Klagegesang – Libanon*, D. Edzard, Ed. (de Gruyter, vol. 6, 1980), pp. 375–379.
8. N. Wasserman, *Akkadian Love Literature of the Third and Second Millennium BCE* (Harrassowitz, vol. 4, 2016).
9. M. Stol, *Women in the Ancient Near East* (de Gruyter, 2016).
10. P. W. Ewald, *Int. J. Paleopathol*. **21**, 178 (2018).
11. B. Mühlemann *et al*., *Proc. Natl. Acad. Sci. U.S.A*. **115**, 7557 (2018).
12. J. Limeres Posse, P. Diz Dios, C. Scully, in *Saliva Protection and Transmissible Diseases* (Academic Press, 2017), pp. 53–92.
13. B. Böck, *The Healing Goddess Gula: Towards an Understanding of Ancient Babylonian Medicine* (Brill, 2014).
14. N. P. Heeßel, *Babylonisch-assyrische Diagnostik* (Ugarit-Verlag, 2000).
15. T. P. Arbøll, *Medicine in Ancient Assur: A Microhistorical Study of the Neo-Assyrian Healer Kişir-Aššur* (Brill, 2021).

ACKNOWLEDGMENTS

The authors acknowledge financial support from the Carlsberg Foundation and the Edubba Foundation.

10.1126/science.adf0512

ENGINEERING

The disappearing boundary between organism and machine

Artificial skin mimics the sensory feedback of biological skin

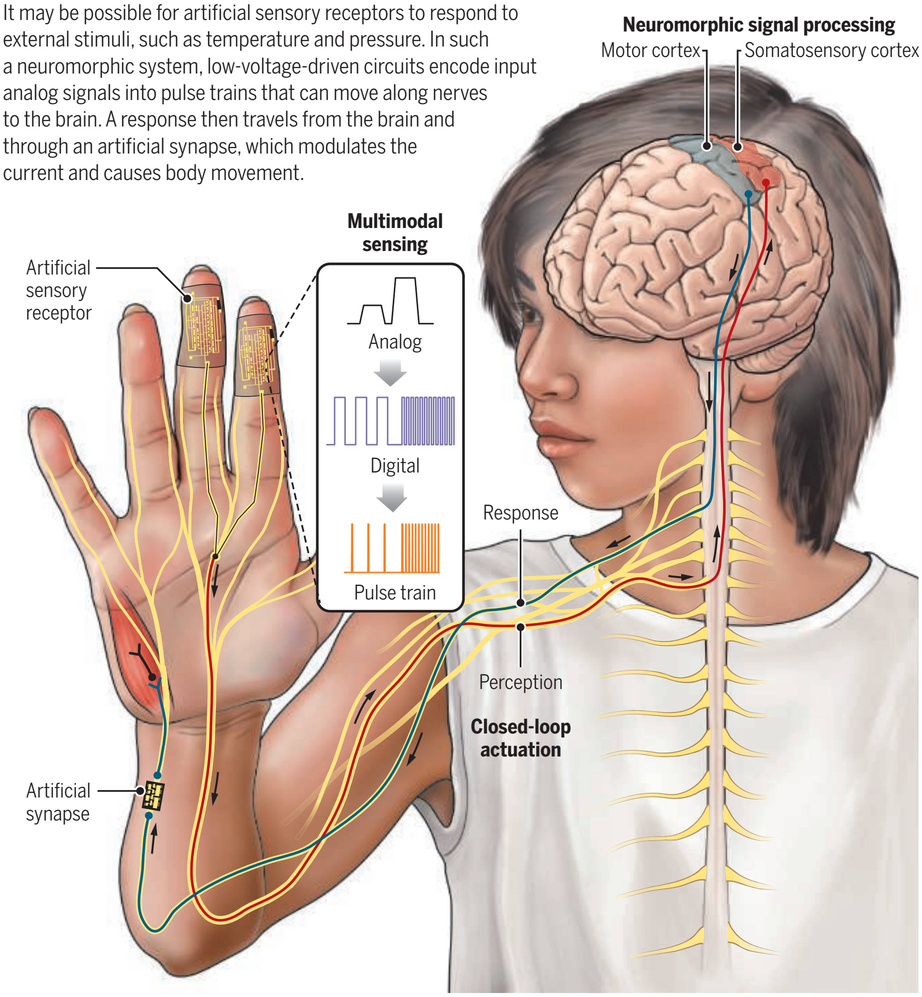
By **Tsuyoshi Sekitani**

Research and development in artificial intelligence (AI) and robotics continue to explore the interface of living organisms with machines. AI-based, motorized prosthetic arms and legs, as well as brain-machine interfaces that connect prosthetics to human brain activity, aim to eventually succeed in fully mimicking human bodily functions and serve as medical solutions for people suffering from illness and injury. However, to more seamlessly integrate mechanical actuators, such as prosthetics, with the human body, the development of a large, elaborate, and stretchable sensor that covers both the body and the mechanical components is necessary (*1*–*6*). Electronic skin (e-skin) would eliminate the boundary between the living body and machine components. On page 735 of this issue, Wang *et al*. (*7*) report the development of a neuromorphic e-skin system that mimics the sensory feedback of biological skin.

Mechanically, the body is a skin-covered sensor and an actuator that converts input energy into an intended physical motion. Sensory organs within the surface of the

An electronic skin that supports a sensorimotor loop

It may be possible for artificial sensory receptors to respond to external stimuli, such as temperature and pressure. In such a neuromorphic system, low-voltage-driven circuits encode input analog signals into pulse trains that can move along nerves to the brain. A response then travels from the brain and through an artificial synapse, which modulates the current and causes body movement.



skin perceive external stimuli, such as pressure, pain, and temperature, with a high level of sensitivity. We can sense danger (stimuli) and immediately move our bodies (reflex response).

The major challenge in developing an e-skin system that fully mimics the functions of living skin is to establish a sensory feedback mechanism. Sensory feedback is the transmission of sensory information perceived by the skin to the brain through nerves. The brain must first process the transmitted sensory signals and then operate the muscles. This sensorimotor nervous system (a closed loop of perception and movement) is essential for realizing natural bodily functions.

Sensor arrays will be crucial for mimicking the skin functions of living organisms; these consist of multiple detectors arranged and aligned in a multidimensional array. By establishing multichannel sensor arrays, it may be possible to mimic the complex sensory functions of human skin. However, generating multiple-row pulse-train signal patterns (*8*) and delivering them to the brain through nerves is a challenge. An organism accomplishes this transmission of sensory information by encoding the input amplitude information with frequency-modulated signals and propagating the sensory information along nerves. In other words, it is simultaneously a signal-digitizing circuit that converts analog signals from sensory receptors into spike-train signal patterns and an artificial synapse that modulates the current amplitude and causes body movement (*9*). To realize a functional human-machine interface suitable for accomplishing daily tasks, an electronic system with skin-like properties that mimics biological functions, such as multimodal sensing, neuromorphic signal processing, and closed-loop actuation (*10*), must be developed, which is more difficult than simply fabricating a sensor array.

Wang *et al.* have developed a neuromorphic e-skin system that fulfills all the essential requirements for integration with the living body. The authors created stretchable field-effect transistors and all-solid-state synaptic transistors that can be driven at a low voltage (0.5 V) as well as technology for their integration (see the figure).

Stretchable organic transistors consist of three-layer, high–permittivity elastomer dielectrics as the core technology, which provide electrical performance—such as low voltage drive, low power consumption, and moderate circuit integration—comparable to that of poly-silicon transistors. By being composed of such stretchable, high-performance organic devices, the e-skin of Wang *et al.* is capable of multimodal perception, the generation of neuromorphological pulse-train signals, and closed-loop actuation, thus mimicking biological sensorimotor loops.

Because all essential components of the e-skin system could reproduce the biological perception-actuation loop, Wang *et al.* tested the utility of this platform in a live rat model. The authors connected the soft e-skin to the rat's somatosensory cortex to reproduce cutaneous sensations. This was expected to elicit a feedback response in the motor cortex. Evoked motor signals would stimulate the sciatic nerve, and, through artificial synapses, the signals would then activate downstream muscles, completing the artificial sensorimotor loop. In the experiment, various magnitudes of pressure applied to the sensor triggered digitized inputs to the somatosensory cortex, which successfully stimulated the firing of neurons in the motor cortex.

The development of neuromorphic e-skin systems is expected to have a substantial effect on a wide range of industrial areas. For example, robots clad in e-skin may be able to feel pain and pressure in the same way that humans do, thus better ensuring the safety of the people they accompany. At present, skin damage and amputation cause a massive disruption in the closed loop of perception and movement, and simple tasks such as grasping an object become challenging for patients. If the closed loop of perception and movement can be reconstructed using the neuromorphic e-skin system, the quality of life of such patients will improve.

Advances in digital technology are uniting information space with physical space. Future research on e-skin should aim for faster sensory responses, finer sensory organs, and a higher density and multimodality of sensor elements for perception. ■

SANKEN, Osaka University, 8-1, Mihogaoka, Ibaraki, Osaka 567-0047, Japan. Email:sekitani@sanken.osaka-u.ac.jp

REFERENCES AND NOTES

1. V. Lumelsky, M. S. Shur, S. Wagner, M. Ding, *Sensitive Skin*, vol. 18 of Selected Topics in Electronics and Systems (World Scientific, 2000).
2. T. Someya *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 9966 (2004).
3. J. A. Rogers, T. Someya, Y. Huang, *Science* **327**, 1603 (2010).
4. T. Someya, Z. Bao, G. G. Malliaras, *Nature* **540**, 379 (2016).
5. A. Chortos, J. Liu, Z. Bao, *Nat. Mater.* **15**, 937 (2016).
6. Y. Kim *et al.*, *Science* **360**, 998 (2018).
7. W. Wang *et al.*, *Science* **380**, 735 (2023).
8. F. Liu *et al.*, *Sci. Robot.* **7**, eabl7344 (2022).
9. R. S. Johansson, J. R. Flanagan, *Nat. Rev. Neurosci.* **10**, 345 (2009).
10. B. C. K. Tee *et al.*, *Science* **350**, 313 (2015).

ACKNOWLEDGMENTS

The author thanks T. Uemura and T. Araki of Osaka University, A. Mori of mojimoji LLC, and research members of the Japan Science and Technology Agency Moonshot Research and Development Program (grant no. JPMJMS2012) for valuable discussions.

10.1126/science.adf0262

ECOLOGY

When the enemy of an enemy is no friend

By **Chris T. Darimont**[1,2] *and* **Ishana Shukla**[1,2]

An ancient proverb suggests that the enemy of one's enemy can be a friend. No such luck, however, for mesocarnivores, such as coyotes and bobcats. For these medium-sized carnivores, humans are a fierce enemy of their chief enemy, the large carnivores that commonly kill them. But such dynamics do not necessarily make humans friends. When embedded within this tripartite relationship, the behavior and associated mortality of mesocarnivores can amplify impacts on their populations. On page 754 of this issue, Prugh *et al.* (*1*) show how such an unexpected outcome can arise. Sharing landscapes in Washington state with large carnivores and humans, mesocarnivores tended to avoid the former by choosing habitats dominated by the latter but paid a high price in mortality. This surprising finding provides new conceptual insights for applied behavioral ecology.

Prugh *et al.* show how interacting dangers posed by humans and large carnivores can affect the behavior and mortality of mesocarnivores. The team collared coyotes (*Canis latrans*), bobcats (*Lynx rufus*), wolves (*Canis lupus*), and cougars (*Puma concolor*). Across two rural regions with considerable human activity, large carnivores strongly avoided human activity, whereas coyotes and bobcats generally shifted their movements away from the larger predators and toward areas with high human influence. This suggested that these mesocarnivores perceived humans as "shields" protecting them from large-carnivore enemies (*2*). This behavioral effect was in contrast to the fitness-related outcome: Across two study areas, the annual risk of mortality from humans was 3 and 3.8 times higher than the risk from large carnivores for coyotes and bobcats, respectively. Such a scenario might represent an ecological trap. These are contexts wherein animals—in apparent error—select habitats in

[1]Department of Geography, University of Victoria, Victoria, BC, Canada. [2]Raincoast Conservation Foundation, Sidney, BC, Canada. Email: darimont@uvic.ca

skin perceive external stimuli, such as pressure, pain, and temperature, with a high level of sensitivity. We can sense danger (stimuli) and immediately move our bodies (reflex response).

The major challenge in developing an e-skin system that fully mimics the functions of living skin is to establish a sensory feedback mechanism. Sensory feedback is the transmission of sensory information perceived by the skin to the brain through nerves. The brain must first process the transmitted sensory signals and then operate the muscles. This sensorimotor nervous system (a closed loop of perception and movement) is essential for realizing natural bodily functions.

Sensor arrays will be crucial for mimicking the skin functions of living organisms; these consist of multiple detectors arranged and aligned in a multidimensional array. By establishing multichannel sensor arrays, it may be possible to mimic the complex sensory functions of human skin. However, generating multiple-row pulse-train signal patterns (*8*) and delivering them to the brain through nerves is a challenge. An organism accomplishes this transmission of sensory information by encoding the input amplitude information with frequency-modulated signals and propagating the sensory information along nerves. In other words, it is simultaneously a signal-digitizing circuit that converts analog signals from sensory receptors into spike-train signal patterns and an artificial synapse that modulates the current amplitude and causes body movement (*9*). To realize a functional human-machine interface suitable for accomplishing daily tasks, an electronic system with skin-like properties that mimics biological functions, such as multimodal sensing, neuromorphic signal processing, and closed-loop actuation (*10*), must be developed, which is more difficult than simply fabricating a sensor array.

Wang *et al.* have developed a neuromorphic e-skin system that fulfills all the essential requirements for integration with the living body. The authors created stretchable field-effect transistors and all-solid-state synaptic transistors that can be driven at a low voltage (0.5 V) as well as technology for their integration (see the figure).

Stretchable organic transistors consist of three-layer, high–permittivity elastomer dielectrics as the core technology, which provide electrical performance—such as low voltage drive, low power consumption, and moderate circuit integration—comparable to that of poly-silicon transistors. By being composed of such stretchable, high-performance organic devices, the e-skin of Wang *et al.* is capable of multimodal perception, the generation of neuromorphological pulse-train signals, and closed-loop actuation, thus mimicking biological sensorimotor loops.

Because all essential components of the e-skin system could reproduce the biological perception-actuation loop, Wang *et al.* tested the utility of this platform in a live rat model. The authors connected the soft e-skin to the rat's somatosensory cortex to reproduce cutaneous sensations. This was expected to elicit a feedback response in the motor cortex. Evoked motor signals would stimulate the sciatic nerve, and, through artificial synapses, the signals would then activate downstream muscles, completing the artificial sensorimotor loop. In the experiment, various magnitudes of pressure applied to the sensor triggered digitized inputs to the somatosensory cortex, which successfully stimulated the firing of neurons in the motor cortex.

The development of neuromorphic e-skin systems is expected to have a substantial effect on a wide range of industrial areas. For example, robots clad in e-skin may be able to feel pain and pressure in the same way that humans do, thus better ensuring the safety of the people they accompany. At present, skin damage and amputation cause a massive disruption in the closed loop of perception and movement, and simple tasks such as grasping an object become challenging for patients. If the closed loop of perception and movement can be reconstructed using the neuromorphic e-skin system, the quality of life of such patients will improve.

Advances in digital technology are uniting information space with physical space. Future research on e-skin should aim for faster sensory responses, finer sensory organs, and a higher density and multimodality of sensor elements for perception. ■

REFERENCES AND NOTES

1. V. Lumelsky, M. S. Shur, S. Wagner, M. Ding, *Sensitive Skin*, vol. 18 of Selected Topics in Electronics and Systems (World Scientific, 2000).
2. T. Someya *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 9966 (2004).
3. J. A. Rogers, T. Someya, Y. Huang, *Science* **327**, 1603 (2010).
4. T. Someya, Z. Bao, G. G. Malliaras, *Nature* **540**, 379 (2016).
5. A. Chortos, J. Liu, Z. Bao, *Nat. Mater.* **15**, 937 (2016).
6. Y. Kim *et al.*, *Science* **360**, 998 (2018).
7. W. Wang *et al.*, *Science* **380**, 735 (2023).
8. F. Liu *et al.*, *Sci. Robot.* **7**, eabl7344 (2022).
9. R. S. Johansson, J. R. Flanagan, *Nat. Rev. Neurosci.* **10**, 345 (2009).
10. B. C. K. Tee *et al.*, *Science* **350**, 313 (2015).

ACKNOWLEDGMENTS

The author thanks T. Uemura and T. Araki of Osaka University, A. Mori of mojimoji LLC, and research members of the Japan Science and Technology Agency Moonshot Research and Development Program (grant no. JPMJMS2012) for valuable discussions.

SANKEN, Osaka University, 8-1, Mihogaoka, Ibaraki, Osaka 567-0047, Japan. Email:sekitani@sanken.osaka-u.ac.jp

10.1126/science.adf0262

ECOLOGY

When the enemy of an enemy is no friend

Some wildlife species mistakenly seek human-inhabited areas to avoid predators

By **Chris T. Darimont**[1,2] *and* **Ishana Shukla**[1,2]

An ancient proverb suggests that the enemy of one's enemy can be a friend. No such luck, however, for mesocarnivores, such as coyotes and bobcats. For these medium-sized carnivores, humans are a fierce enemy of their chief enemy, the large carnivores that commonly kill them. But such dynamics do not necessarily make humans friends. When embedded within this tripartite relationship, the behavior and associated mortality of mesocarnivores can amplify impacts on their populations. On page 754 of this issue, Prugh *et al.* (*1*) show how such an unexpected outcome can arise. Sharing landscapes in Washington state with large carnivores and humans, mesocarnivores tended to avoid the former by choosing habitats dominated by the latter but paid a high price in mortality. This surprising finding provides new conceptual insights for applied behavioral ecology.

Prugh *et al.* show how interacting dangers posed by humans and large carnivores can affect the behavior and mortality of mesocarnivores. The team collared coyotes (*Canis latrans*), bobcats (*Lynx rufus*), wolves (*Canis lupus*), and cougars (*Puma concolor*). Across two rural regions with considerable human activity, large carnivores strongly avoided human activity, whereas coyotes and bobcats generally shifted their movements away from the larger predators and toward areas with high human influence. This suggested that these mesocarnivores perceived humans as "shields" protecting them from large-carnivore enemies (*2*). This behavioral effect was in contrast to the fitness-related

[1]Department of Geography, University of Victoria, Victoria, BC, Canada. [2]Raincoast Conservation Foundation, Sidney, BC, Canada. Email: darimont@uvic.ca

outcome: Across two study areas, the annual risk of mortality from humans was 3 and 3.8 times higher than the risk from large carnivores for coyotes and bobcats, respectively. Such a scenario might represent an ecological trap. These are contexts wherein animals—in apparent error—select habitats in which fitness is lower compared with other available habitats (*3*).

Prugh *et al.* referred to this scenario as "the paradox of the lethal human shield." Human shields are a form of commensal "protective association" in which one species benefits from reduced predation rate when near another, protective species (*4*). In the case of human shields, some species (or members within species) seek association with humans as an apparent means to protect them from greater perceived risk. For example, in the Yellowstone system, moose (*Alces alces*) shifted calving locations toward roads, presumably to avoid predation by road-averse grizzly bears (*Ursus arctos*). Fitness outcomes explain how this general behavior could evolve. In another example, Steyaert *et al.* (*5*) showed how female grizzlies with surviving litters were more likely to use humans as protective spatial associates compared to mothers that lost cubs. The latter group avoided humans and co-occurred more often with infanticidal males, which also avoided humans. In striking contrast, Prugh *et al.* reveal the paradoxical scenario in which the result (higher mortality) appears to contradict the fitness-related premise of the human shield argument. This paradox opens the door to the possibility of an ecological trap. Only additional testing related to habitat selection and fitness measures in defined areas with and without human shields, however, could provide such evidence (*3*). Nonetheless, the findings of Prugh *et al.* identify an important scenario wherein wildlife species poorly evaluate risks from humans, which might be more common than appreciated.

Why would wildlife move toward humans, given the enormous risk? After all, humans kill wildlife—especially medium-sized and large carnivores—at far higher rates compared to other predators (*6*). Indeed, many species across the globe tend to avoid areas dominated by human activity (*7*). Moreover, experimental evidence suggests that carnivores and herbivores can fear humans more than they fear other carnivores (*8*, *9*). Given these abundant and compelling data, why might some wildlife behave so unexpectedly in space use? Although not evaluated, the coyotes and bobcats could have sought to limit risk with more nocturnal behavior near human activity (*10*). Moreover, Prugh *et al.* point out that excess mortality might be offset by fitness-related benefits, such as food resources, in human-dominated areas. The authors also suggest that these mesocarnivores simply might not be competent at gauging risk from humans.

A microevolutionary process might underlie poor risk judgment and explain why human shields—and other human-dominated contexts—might become ecological traps. In some cases, humans might be too successful a predator for prey to coevolve effective antipredatory behavior. At least in the context of morphological adaptations, unsuccessful predation is a necessary condition for adaptation (*11*). The failed attempts by predators to secure prey allows surviving individuals to pass on traits beneficial to escaping search, pursuit, or manipulation stages of predation (*12*). In the context of harvesting, species of shoreline marine invertebrates have not adapted morphologically to human harvesting, instead likely relying on spatial refuges from harvesters (*13*). Perhaps such avoidance is not possible for some terrestrial species. The apparent evolutionary mismatch by mesocarnivores hints at the possible limits to behavioral adaptation of prey imposed by humans, ultraefficient and omnipresent predators on land. Although data are scarce, failure rates by contemporary hunters, trappers, and other terrestrial harvesters are likely much lower than those by large carnivores. Moreover, contemporary humans can harvest almost anywhere. As a result, perhaps too few survivors pass on behavioral tendencies that equip their descendants to competently evaluate risk during the human predator's ever-present search phase to select for fear responses commensurate with the extraordinary risk.

Despite uncertainty among underlying behavioral and evolutionary processes, the study by Prugh *et al.* highlights the ecological implications of human influence on relationships among multiple trophic levels. High levels of human disturbance can create an additional upper level, imposing strong and varied top-down effects through both increased mortality and trait-mediated indirect interactions (*14*). Future research could examine whether misalignment between behavior and mortality of ecologically influential species is more widespread. A better understanding of failure rates of human predators could also yield new insight into how animals adapt to a world in which high human-caused mortality of wildlife becomes a defining feature of the Anthropocene. ■



When coyotes move to regions dominated by humans to escape large predators, their mortality can increase.

REFERENCES AND NOTES

1. L. R. Prugh *et al.*, *Science* **380**, 754 (2023).
2. J. Berger, *Biol. Lett*. **3**, 620 (2007).
3. R. Hale, S. E. Swearer, *Proc. Biol. Sci*. **283**, 20152647 (2016).
4. J. L. Quinn, M. Ueta, *Ibis* **150**, 146 (2008).
5. S. M. J. G. Steyaert *et al.*, *Proc. Biol. Sci*. **283**, 20160906 (2016).
6. C. T. Darimont, C. H. Fox, H. M. Bryan, T. E. Reimchen, *Science* **349**, 858 (2015).
7. M. A. Tucker *et al.*, *Science* **359**, 466 (2018).
8. K. M. Gaynor, C. E. Hojnowski, N. H. Carter, J. S. Brashares, *Science* **360**, 1232 (2018).
9. J. P. Suraci, J. A. Smith, M. Clinchy, L. Y. Zanette, C. C. Wilmers, *Sci. Rep*. **9**, 12214 (2019).
10. D. A. Crawford, L. M. Conner, M. Clinchy, L. Y. Zanette, M. J. Cherry, *Oecologia* **198**, 91 (2022).
11. G. J. Vermeij, *Am. Nat*. **120**, 701 (1982).
12. T. E. Reimchen, M. A. Bell, S. A. Foster, in *The Evolutionary Biology of the Threespine Stickleback*, M. A. Bell, S. A. Foster, Eds. (Oxford Univ. Press, 1994), pp. 240–276.
13. G. J. Vermeij, *Evolution* **66**, 2007 (2012).
14. M. W. Wilson *et al.*, *Ecol. Lett*. **23**, 1522 (2020).

ACKNOWLEDGMENTS

The authors thank T. Reimchen and J. Suraci for helpful feedback on a draft.

10.1126/science.adh9166

CLIMATOLOGY

Global loss of lake water storage

Drying trends are prevalent worldwide

By **Sarah W. Cooley**

Lakes sustain a wide variety of ecosystems and provide vital water for agriculture, hydropower, and direct human consumption. Often characterized as "sentinels of climate change" (*1*), lakes integrate multiple basin-scale climatic processes including precipitation, runoff, and evapotranspiration. The amount of water stored in lakes therefore reflects both short- and long-term climate fluctuations. However, attributing changes in lake water storage to climate is complex, because direct human activities such as reservoir management, water withdrawals, and land-use change also affect lake water storage. On page 743 of this issue, Yao *et al*. (*2*) present a dataset of decadal-scale trends in lake water storage from 1992 to 2020 and attribute them to human activities and climatic patterns.

Lake water storage fluctuations are primarily driven by the balance between precipitation and evapotranspiration in a lake's watershed (*3*). As the climate warms, evapotranspiration is broadly increasing owing to increasing air temperatures and changing surface energy balance (*4*). The effect of climate warming on precipitation is less globally consistent, but in some regions (e.g., the Arctic, the Upper Mississippi), precipitation increases are expected (*5*). As a result, climate change can lead to both net water gain and net water loss in lakes. The "dry gets drier, wet gets wetter" paradigm is often used to summarize these patterns, though this can break down locally, and local characteristics can complicate the interpretation of global-scale trends. For example, changes in other forms of water storage can also influence long-term patterns, such as on the Tibetan Plateau, where increasing glacier melt has led to net gains in water volume (*6*).

This complexity in understanding long-term, climatic drivers of lake water storage is further magnified by direct human activities, which can affect lake water storage in both natural lakes and human-managed reservoirs. Human diversion of water for irrigation, for example, can lead to declining water inputs to lakes, thus causing substantial water loss (*7*). In human-managed reservoirs, an additional threat to water storage is sedimentation, whereby the presence and operation of a dam traps sediment that would otherwise travel downstream. Over time, this sediment can accumulate and reduce a reservoir's storage capacity (*8*). Conversely, construction of new dams and reservoirs can increase lake water storage. Although dam construction has declined in North America and Europe, in many river basins in Asia, Africa, and South America, extensive dam construction has altered lake water storage patterns (*9*).

By building a water storage variability dataset for 1980 of the world's largest lakes, Yao *et al*. identified a widespread net decline in lake water storage (21.72 ± 2.11 Gt year^{-1}), finding that drying trends are prevalent worldwide. The authors attribute most of this trend in natural lakes to increasing air temperature and evaporative demand, and to direct human activities. These trends are global but tend to be stronger in arid climates. Roughly one-third of the global drying trend in natural lakes is offset by lake water storage increases, predominantly caused by increases in precipitation and runoff. The analysis also reveals a declining trend in reservoir water storage in both arid and humid regions owing to sedimentation, though the recent boom in reservoir construction, primarily in humid regions, has led to a net increase in reservoir water storage.

"...nearly one-quarter of the world's population lives in a basin with a large, drying lake."

These insights into lake water storage trends are enabled by advances in cloud computing, satellite altimeters (*10*), and reservoir datasets (*11*). Yao *et al*. produced their dataset by fusing time-varying water areas mapped using Landsat satellite imagery with water surface elevation measurements from satellite altimeters to produce a 30-year dataset of lake volume variability. Previous work either focused on seasonal variability (*12*) or was limited by gaps in the satellite data record (*13*). Improvements in satellite technologies and data availability may better elucidate global patterns in storage variability. For example, NASA's Surface Water Ocean Topography mission will soon provide regular water level observations for millions of lakes globally (*14*), potentially enabling expansion of this analysis to include many more water bodies, including small lakes.

The global-scale decline in lake water storage in arid regions falls within the "dry gets drier" framework of climate impacts on the water cycle, but the drying trends in humid regions complicate the "wet gets wetter" part. Although this result counters previous work on long-term lake storage trends (*13*), it agrees with climate-modeling analyses challenging the global applicability of the "dry gets drier, wet gets wetter" framework (*15*). This lack of storage increases in natural lakes in humid regions, which is attributed to increasing evapotranspiration, emphasizes that we should not expect greater water availability in humid basins to offset water losses in arid basins.

Yao *et al*. estimated that nearly one-quarter of the world's population lives in a basin with a large, drying lake. Considering the importance of these lakes for ecosystem services, water supply, irrigation, and/or hydropower, the potential consequences of lake drying are both locally and globally important. Future work may combine observed drying trends with climate modeling to better constrain these changes at local to global scales. ■

Department of Geography, University of Oregon, Eugene, OR, USA. Email: scooley2@uoregon.edu

REFERENCES AND NOTES

1. R. Adrian *et al*., *Limnol. Oceanogr*. **54**, 2283 (2009).
2. F. Yao *et al*., *Science* **380**, 743 (2023).
3. R. I. Woolway *et al*., *Nat. Rev. Earth Environ*. **1**, 388 (2020).
4. M. L. Roderick, F. Sun, W. H. Lim, G. D. Farquhar, *Hydrol. Earth Syst. Sci*. **18**, 1575 (2014).
5. K. Trenberth, *Clim. Res*. **47**, 123 (2011).
6. G. Zhang *et al*., *Earth Sci. Rev*. **208**, 103269 (2020).
7. P. Micklin, *Lakes Reservoirs: Res. Manage*. **15**, 193 (2010).
8. A. J. Schleiss, M. J. Franca, C. Juez, G. De Cesare, *J. Hydraul. Res*. **54**, 595 (2016).
9. C. Zarfl, A. E. Lumsdon, J. Berlekamp, L. Tydecks, K. Tockner, *Aquat. Sci*. **77**, 161 (2015).
10. L. Magruder, T. Neumann, N. Kurtz, *Earth Space Sci*. **8**, e2020EA001555 (2021).
11. J. Wang *et al*., *Earth Syst. Sci. Data* **14**, 1869 (2022).
12. S. W. Cooley, J. C. Ryan, L. C. Smith, *Nature* **591**, 78 (2021).
13. S. Luo *et al*., *Geophys Res Lett*. **49**, e2021GL096676 (2022).
14. S. Biancamaria, D. P. Lettenmaier, T. M. Pavelsky, *Surv. Geophys*. **37**, 307 (2016).
15. M. P. Byrne, P. A. O'Gorman, *J. Clim*. **28**, 8078 (2015).

10.1126/science.adi0992

POLICY FORUM

CONSERVATION

Braiding Indigenous rights and endangered species law

Recovery targets fall short of culturally meaningful abundance

By **Clayton T. Lamb[1,2], Roland Willson[3], Allyson K. Menzies[4], Naomi Owens-Beek[5], Michael Price[6], Scott McNay[7], Sarah P. Otto[8,9], Mateen Hessami[2], Jesse N. Popp[4], Mark Hebblewhite[10], Adam T. Ford[1,2]**

Endangered species laws effectively prevent species extinction but fall short in restoring abundance for culturally important species. Legal agreements between Indigenous peoples and countries recognize the importance of abundant, culturally important species that disproportionately contribute to peoples' food, material, spirituality, and sense of place (*1*). Despite this, recovery targets under endangered species laws do not account for such abundance, instead targeting minimum viable population (MVP) sizes that leave many species in a state of reduced abundance compared with their historical baselines. Using three keystone species in North America—caribou, bison, and salmon—we explore the implications of the gap between culturally meaningful abundance and minimum viable populations and argue for the need to establish recovery targets and processes that restore abundance beyond MVPs. Braiding endangered species law and Indigenous rights will help countries uphold the rights of Indigenous peoples, prevent species extinction, and ultimately provide benefits to society at large.

Under endangered species laws in Canada and the United States, recovery targets are left vague or are based on a MVP. In Canada, the Species at Risk Act (SARA) focuses on risks of extinction and does not explicitly define recovery, with a recent 2020 SARA policy document interpreting recovery in terms of reducing the risk of extinction or extirpation. In the United States, the focus of the Endangered Species Act (ESA) is on meeting targets of a MVP, that is, a species abundance that will enable population persistence with minimal human intervention. Such risk- and MVP-based approaches generate modest recovery targets, tending to simply maintain populations at low levels in most recovery documents (*2*) [see supplementary materials (SM) for more detail on each act].

Critically, neither SARA nor ESA address how people interact with the species through harvest. Harvest reflects time-honored relationships that support food security, ceremonial practices, or other hallmarks of culture (*3*), yet harvested species are often excluded from listing under SARA because of socioeconomic concerns (*4*). A culturally meaningful recovery target for such species may require a greater abundance and/or different distribution than those prescribed by risk- or MVP-based approaches. Culturally meaningful recovery also requires more-inclusive policies to center Indigenous perspectives and people in the design and implementation of restorative actions (*5*–*7*).

Here, we describe three recovery efforts that demonstrate continued inequities in biodiversity conservation policies. We highlight the need to reconsider recovery targets for culturally important, harvested species in national endangered species laws and policies. We focus on three high-profile species in North America—caribou, bison, and salmon—which have formed central aspects of Indigenous peoples' diet, culture, and seasonal movements since time immemorial. In each case, the decline of these species impeded Indigenous peoples from carrying out cultural practices and exercising food sovereignty. Each of these three culturally important species has since shown some level of recovery, and we highlight how these recoveries—often considered conservation victories—remain distant from culturally meaningful levels of recovery. This mismatch is partly due to a lack of formal legislation supporting culturally meaningful recovery targets (see the first figure and SM section 1). The species highlighted here are emblematic of the many culturally important species that are, at present, in a state of diminished abundance across the globe (*1*).

CARIBOU

Woodland caribou (*Rangifer tarandus*) have long been a primary food source for northern Indigenous peoples in North America. Caribou populations have declined extensively in the past century (see the second figure), especially in the southern portion of their range. Eleven of 38 southern mountain caribou subpopulations are extirpated, and the overall population has declined by more than 40% during the past 20 years, as observed using Western monitoring techniques (*5*). Indigenous Knowledge, whose relevance and value are increasingly being recognized by colonial governments and agencies, has been providing an invaluable historical baseline of abundance and harvest levels since well before Western science was engaged in species recovery. For example, in British Columbia, Canada, Elders from the Treaty No. 8 adherent West Moberly First Nations said that the Klinse-Za caribou subpopulation was once as abundant as bugs on the landscape, yet, by 2013, there were only 38 animals left (*5*). Facing a decline in caribou, West Moberly leadership and Elders imposed a moratorium on caribou harvest in 1970 that is still in effect today. Indeed, West Moberly First Nations sensed the endangerment of caribou well before colonial governments, who continued to permit extensive natural resource extraction in the heart of the Klinse-Za caribou habitat after West Moberly's cessation of hunting. The continued authorization of resource extraction reduces the abundance of caribou and causes extirpation, infringing on constitutionally protected Indigenous rights to sustain a culturally meaningful way of life (*8*).

Indigenous-led recovery efforts by West Moberly First Nations and Saulteau First Nations to recover the threatened Klinse-Za caribou have more than tripled caribou abundance in 8 years (*5*). Averting the looming extirpation of these caribou is an undeniable conservation success, yet their abundance remains below a level at which First Nations can participate in a culturally meaningful harvest. In 2022, there were 114 Klinse-Za caribou, an abundance that met a recovery target of >100 set by the Canadian government under the MVP-based approach. However, 114 caribou would provide only about three caribou annually for a sustainable Indigenous harvest, which does not meet historical levels of use by the community (see SM for harvest calculations). A West Moberly Elder's wish was to "eat caribou

[1]Department of Biology, University of British Columbia, Kelowna, BC, Canada. [2]Wildlife Science Center—Biodiversity Pathways, University of British Columbia, Kelowna, BC, Canada. [3]West Moberly First Nations, Moberly Lake, BC, Canada. [4]School of Environmental Sciences, University of Guelph, Guelph, ON, Canada. [5]Saulteau First Nations, Moberly Lake, BC, Canada. [6]Earth to Ocean Research Group, Department of Biological Sciences, Simon Fraser University, Burnaby, BC, Canada. [7]Wildlife Infometrics, Mackenzie, BC, Canada. [8]Department of Zoology, University of British Columbia, Vancouver, BC, Canada. [9]Biodiversity Research Centre, University of British Columbia, Vancouver, BC, Canada. [10]Wildlife Biology Program, University of Montana, Missoula, MT, USA. Email: ctlamb@ualberta.ca

before I die," which could be translated to a baseline cultural recovery target. If caribou are to be meaningfully harvested again, there should be enough for each community member to have some meaningful level of cultural engagement and food security met by caribou. For example, providing just one meal per year for each of the 1600 West Moberly and Saulteau First Nations members would require about six caribou, which could be annually harvested from a population of about 200 caribou. Providing 15 meals per person per year would require a population of about 3000 caribou, which is more reflective of the historic "bugs on the landscape" abundance. The discrepancy between the existing 114 caribou and the potential for >3000 caribou is a measurable gap in Western and Indigenous perspectives on recovery and reconciliation.

AMERICAN BISON

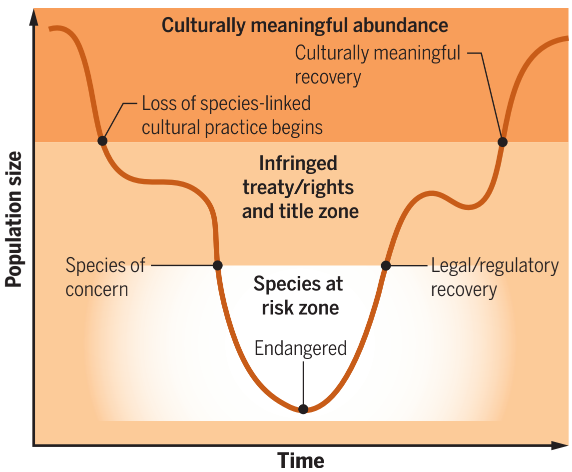
Before colonization, American bison (*Bison bison*) numbered 30 million to 60 million across North America (*6*). Many Indigenous peoples were deeply dependent on this once-abundant species, which ranged from Alaska to Mexico (see SM). By the turn of the 20th century, however, the great bison herds had been slaughtered down to only a few hundred animals, in part driven by explicit policies of cultural genocide (see the second figure). Such pronounced bison declines caused starvation, infighting, and the erosion of Indigenous culture (*9*).

The northern subspecies of American bison, wood bison (*Bison bison athabascae*), were listed as endangered under the ESA in 1979 and classified as endangered by the Committee on the Status of Endangered Wildlife in Canada (COSEWIC) in 1978 (predating SARA). Recovery actions included establishing new wood bison populations, which led to an increased abundance of nearly 10,000 individuals within Canada by 2013. As a result, COSEWIC downlisted them to threatened, with at least five free-ranging populations of 1000 bison as the MVP goal. Despite this example of MVP recovery, some Indigenous peoples are often still prevented from harvesting bison. For example, although limited harvest does occur across the wood bison range, many Indigenous peoples in northern Canada were excluded from hunting within Wood Buffalo National Park despite the recovery of the population to a level that could sustain a harvest. Similarly, subsistence hunting rights remain curtailed for the wood bison herd of Aishihik, where Indigenous peoples of the Yukon are not permitted these rights because the animals originated from a transplant. Thus, wood bison recovery highlights the continued mismatch between MVP recovery and the culturally meaningful recovery of wood bison.

The precipitous decline of the southern bison subspecies, plains bison (*Bison bison bison*), occurred well before SARA or ESA came into effect (in 2002 and 1973, respectively). Nevertheless, wild plains bison are still clearly endangered. Today, the species remains at <1% of its historic abundance and occupies a markedly reduced range (see the second figure). Despite early recovery efforts, wild, free-ranging plains bison populations represent only 10% of the current abundance of plains bison; the remaining 90% are privately owned (*6*). Yet, plains bison remain unlisted in both Canada and the United States despite clear scientific recommendations to do so (*10*).

Diagram of species abundance over time

Violation of Indigenous rights, such as harvest, occurs at a much higher abundance than the focus of present endangered species laws, which is around achieving minimum viable populations.



GRAPHIC: D. AN-PHAM/*SCIENCE* BASED ON FUSE CONSULTING

Recently, an inspiring example of Indigenous leadership in plains bison restoration began unfolding. On 23 September 2014, 13 First Nations and Native American tribes signed the first intertribal treaty in 150 years, the Buffalo Treaty, which focused on the ecological and cultural recovery of plains bison (*6*). Supported in part by the Buffalo Treaty, Banff National Park, Canada, initiated a plains bison restoration program in 2017. The long-term reintroduction goal in Banff is to include co-management of a culturally meaningful bison harvest with Buffalo Treaty signatories. The contemporary successes of bison restoration have increased the likelihood of bison recovery under MVP criteria. But bison recovery will remain incomplete until peoples' cultural connection with bison—including, perhaps, a prominent role in diet—is restored across broader landscapes.

PACIFIC SALMON

Millions of salmon (*Oncorhynchus* spp.) return annually to rivers across western North America, providing sustenance for people, wildlife, and ecosystems. Indigenous peoples in the Pacific Northwest often refer to themselves as "salmon people," signifying their deeply rooted cultural connections with salmon (*7*).

The distribution and abundance of salmon have decreased over the past century through the effects of human activities. In the Columbia Basin region of the United States, salmon abundance has declined by ~75%; an estimated 7.5 million to 16 million salmon returned annually to the Columbia Basin before the 20th century, and now only 1 million to 4 million return (see the second figure) (*11*). The most commercially valuable and culturally important salmon species in Canada, sockeye salmon (*O. nerka*), declined in wild abundance by 69% over the past century in the country's second-largest salmon watershed, the Skeena River (*12*). Salmon harvests by Indigenous communities in Canada have declined by more than 80% in the past 50 to 70 years (*7*), with some First Nations implementing self-imposed harvest bans (*12*).

Salmon recovery is demonstrably underserved by existing endangered species legislation. No salmon population has been listed in Canada under SARA, and although many have been listed in the United States under the ESA, abundance remains a fraction of historic levels. Given the lack of formal protection, several Indigenous-led recovery plans for salmon have recently been developed. For example, after having endured ~60 years of diminished sockeye salmon returns, the Wet'suwet'en Nation on the west coast of Canada have implemented a rebuilding plan with an abundance target set to provide for community and ecosystem needs. However, ongoing commercial fisheries and industrial development projects undermine salmon recovery efforts. Thus, there remains a need for increased recognition of Indigenous rights that support the protection of diminished populations beyond endangered species legislation.

INDIGENOUS RIGHTS

Although international agreements and domestic laws compel governments to recover endangered species, colonial governments are also obligated to honor the legal treaty and constitutional rights of Indigenous

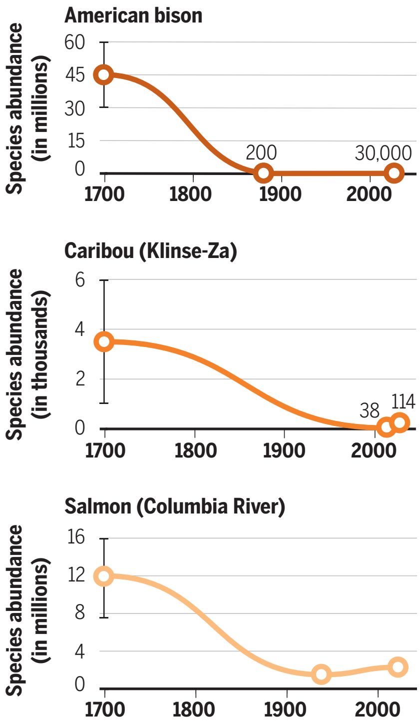
peoples, including rights to fish, hunt, and trap. In some cases, culturally important species are at the center of the interaction between Indigenous and non-Indigenous governments. For example, during negotiations of Treaty No. 8 in 1899, Canada promised Indigenous peoples in Treaty 8, which encompasses nearly 10% of Canada, that they "would be as free to hunt and fish after treaty as they would if they never entered into it" (*13*). A century and a half of colonization on these lands has substantially impeded the ability of Treaty 8 First Nations to hunt and fish as they once did (*8*). Treaty infringement was recently affirmed in the 2021 court case *Blueberry River First Nations (Yahey)* v. *Province of British Columbia*, which concluded that the Province of British Columbia had breached Treaty No. 8 and infringed the rights of Blueberry River First Nations. This was due to the cumulative impacts of rampant resource authorizations and development, which affected culturally important species such as caribou and moose.

We see multiple paths forward that could support increased recognition of legal obligations to Indigenous peoples and recover species to culturally meaningful abundances. A first path includes defining more ambitious recovery targets while still working within the confines of endangered species laws. Recovery plans for culturally important species could propose MVP targets as only a preliminary step toward full recovery. For species with abundances greater than a MVP, recovery status could be assessed against a new global standard, the International Union for Conservation of Nature (IUCN) Green List of species that assesses the degree a species has recovered, which complements the Red List that has been measuring species' risk of extinction since 1964. The Green List sets out ambitious recovery targets—such as "full recovery," which is defined by restoring historic abundance, distribution, and ecological function—and measures species' progress toward these targets (*14*). We recommend that full recovery also include abundance targets that support food security, materials, and cultural relationships that rely on these animals. The 2016 wood bison recovery strategy provides a rare example of abundance targets to support Indigenous rights and particularly a culturally meaningful harvest (*15*). Such culturally meaningful recovery targets will likely be of similar magnitude to historic abundance but must also accommodate the evolving practices, cultures, and communities of Indigenous peoples.

We acknowledge that full recovery will remain challenging for some species, such as plains bison, because of decreases in their historic habitat due to agriculture, urbanization, transportation infrastructure, and resource extraction. In such cases, a modified recovery target based on remaining or restorable habitat may be required. In all cases, culturally meaningful recovery targets must be codeveloped with Indigenous peoples and reflect their present and desired future relationships with a species and the land.

Culturally meaningful North American species decline

Abundance estimates are for a species or a focal population with sufficient data to characterize a broader regional trend. Estimated abundances before colonization and large-scale industrial impacts are shown at year 1700. The lowest recorded population estimate after colonization is shown between 1900 and 2013; the most current estimate (2022) is shown on the right. Interpolated lines between estimates are for visualization purposes only (see supplementary materials for details on data and ranges).



A second path includes legal enforcement of Indigenous rights. Consider the *Yahey* decision that extends beyond legally endangered species and includes many species whose populations are below culturally meaningful abundance thresholds. Here, litigation triggered mechanisms of recovery based on protection of Indigenous rights (see the first figure), which, to date, have included initial reparations of $65 million to Blueberry River First Nations in 2021 and, in January 2023, a $200 million restoration fund, an $87.5 million financial package, more than 650,000 ha of land protections, and obligations related to land-use planning and wildlife management to begin healing the land, increasing the abundance of harvested species, and protecting the Nations' cultural way of life. Recovering species abundance to culturally meaningful levels would satisfy important aspects of presently infringed treaties between Indigenous peoples and governments.

We outline two potential solutions to recover culturally meaningful abundance, but it could be the case that entirely new laws are needed to support such recovery in some countries. We see the harmonization of biodiversity agreements with international agreements such as the United Nations Declaration on the Rights of Indigenous Peoples (UNDRIP) as an opportunity to provide the foundation for recultivating available, accessible, and adequate food, with strong nutritional, cultural, and spiritual connections to a single species or entire ecosystems. This foundation would support creation of new laws to specifically address Indigenous rights violations and wildlife abundance shortfalls. Given that a fundamental reason for conserving species relates to human values of biodiversity, nature, and a responsibility to all life, restoring the very connections that propel recovery will serve to make efforts more successful while protecting critical relationships between people and the land. ■

REFERENCES AND NOTES

1. V. Reyes-García *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2217303120 (2023).
2. K. A. Pawluk, C. H. Fox, C. N. Service, E. H. Stredulinsky, H. M. Bryan, *PLOS ONE* **14**, e0224021 (2019).
3. P. Priadka, B. Moses, C. Kozmik, S. Kell, J. N. Popp, *Ecol. Soc.* **27**, 30 (2022).
4. C. S. Findlay, S. Elgie, B. Giles, L. Burr, *Conserv. Biol.* **23**, 1609 (2009).
5. C. T. Lamb *et al.*, *Ecol. Appl.* **32**, e2581 (2022).
6. H. Shamon *et al.*, *Front. Ecol. Evol.* **10**, 826282 (2022).
7. A. J. Reid *et al.*, *Fish Fish.* **22**, 243 (2021).
8. B. R. Muir, A. L. Booth, *Environ. Dev. Sustain.* **14**, 455 (2012).
9. J. Daschuk, *Clearing the Plains: Disease, Politics of Starvation, and the Loss of Indigenous Life* (Univ. Regina Press, 2019).
10. C. H. Freese *et al.*, *Biol. Conserv.* **136**, 175 (2007).
11. C. L. Smith, "Salmon abundance and diversity in Oregon Are we making progress?" (M/A-21, Oregon State University, 2014).
12. M. H. H. Price, J. W. Moore, B. M. Connors, K. L. Wilson, J. D. Reynolds, *J. Appl. Ecol.* **58**, 1477 (2021).
13. Government of Canada, "Report of the Commissioners for Treaty No. 8." (1899); https://www.rcaanc-cirnac.gc.ca/eng/1100100028813/1581293624572.
14. M. K. Grace *et al.*, *Conserv. Biol.* **35**, 1833 (2021).
15. Environment and Climate Change Canada, "Recovery strategy for the wood bison (*Bison bison athabascae*) in Canada," Species at Risk Act Recovery Strategy Series (Environment and Climate Change Canada).

ACKNOWLEDGMENTS

We thank the Mitacs Accelerate program, the Nîkanêse Wah tzee Stewardship Society, a Liber Ero Fellowship, National Science Foundation grant no. 2127272, and the Canadian Mountain Network for support. This work benefited from review by J. Lam of Sage Legal and M. Olyszynski at the University of Calgary's Faculty of Law. D. Ranglack provided advice on the present and historic abundances and distributions of bison. C.T.L. and R.W. contributed equally to this work.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adg9830

10.1126/science.adg9830



Globes made in the 18th and early 19th centu
offer various representations of Mars' surface.

PHOTO: © WHIPPLE MUSEUM OF THE HISTORY OF SCIENCE, CAMBRIDGE (WH.1268; WH.6622; WH.6625; WH.6238; WH.6211)

PLANETARY SCIENCE

Moving beyond "Why Mars?"

A historian probes the cultural contexts of our enduring fascination with the red planet

By **Robert Markley**

Ever since the Italian astronomer Giovanni Schiaparelli announced in 1878 that he had observed *canali* (channels or canals) crisscrossing its surface, Mars has been the subject of thousands of scientific articles and hundreds of scholarly books and has served as the setting for science-fiction novels and movies almost too numerous to count. In different ways, much of the scientific literature on the planet poses a version of the question "Why Mars?" and then explains, from a variety of perspectives, humankind's millennia-long fascination with the red planet. In *For the Love of Mars*, Matthew Shindell takes a different approach by focusing on what he calls the "who?" questions: Who has studied Mars, who reads all these books and articles about Mars, and who ultimately will be able to travel to Mars? In examining these questions, Shindell crafts a human-centered history of an alien world.

For the Love of Mars offers a readable history of speculation and exploration about Mars, beginning with the skywatchers of ancient Mayan, Chinese, and Babylonian civilizations and ending with the scientists and engineers who work on current robotic missions. Shindell situates these observers within their cultural and socioeconomic contexts, whether the courts of the Han dynasty or the warrens of the Jet Propulsion Laboratory (JPL). Every generation, he demonstrates, recasts Mars in its own historical idioms and its own understandings of humankind and the cosmos.



For the Love of Mars: A Human History of the Red Planet
Matthew Shindell
University of Chicago Press, 2023. 248 pp.

In the era of telescopic observation (circa 1800 to 1964), astronomers described Mars in terms of its earthly analogs. The canal controversy, triggered by Schiaparelli and turned into front-page news by the American businessman and astronomer Percival Lowell, was fueled by widespread fascination with the herculean construction projects of the Suez and Panama canals. Although thousands of workers died building these projects, their deaths could be justified or shrugged aside by the expansionist logic of European imperialism and American manifest destiny. When H. G. Wells skewered this logic in *The War of the Worlds*, his martians ushered in an era characterized by daydreams of colonizing the red planet and nightmares of seeing its forbidding surface as a harbinger of ecological devastation on Earth.

Like prior histories of martian exploration, many of which can be found in the book's extensive bibliography, Shindell moves chronologically from ancient civilizations (chapter 1), through the astrological imperatives of the Middles Ages and the early modern era (chapter 2), the early years of telescopic observation (chapter 3), the canal controversy, and the dawning recognition in the mid-20th century that life on Mars was, at best, a hypothesis, not a scientific fact (chapter 4). His true interest, however, lies in the era of robotic exploration.

If the rhetoric of conquest and colonization in the early 20th century encouraged scientists and science-fiction writers to envision Mars as a new frontier to exploit, then the rhetoric of Cold War rivalry and the space race galvanized the US and USSR to pour billions into sending landers and then rovers to explore the planet's surface (chapter 5). Shindell's final chapter brings readers to the 21st century, describing new generations of Mars obsessives who pore over terabytes of data as eagerly as astrologers 500 years ago pored over their star charts.

Shindell is persuasive in explaining how, in the wake of the Mariner and Viking missions, the justifications for "Why Mars?" were redefined in terms of institutional and technological histories, notably those of NASA and JPL. Drawing on a wide range of sources, Shindell shows how the quest for life on Mars became a central talking point for NASA in the late 20th century, largely as a rationale for the agency's wish list of ambitious missions. Although he begins the book with accounts of the 2020 launches of Mars spacecraft by the United Arab Emirates and China, the "who?" questions that Shindell asks are largely focused on US perspectives.

For the Love of Mars is very much a 21st-century book in its emphasis on how scientific communities have formed, expanded, and diversified over the past 60 years. These collaborations among engineers, geologists, and roboticists have garnered new interest with the rise of the internet and the thousands of photographs of the martian surface returned by a quarter-century of rovers: Sojourner, Spirit, Opportunity, Curiosity, and Perseverance.

Mars has become a recognizable, if still alien, landscape. For many scientists who make up the extended networks of earthbound, would-be martians, the ultimate goal of current robotic exploration and future sample-return missions is human expeditions to Mars. Elon Musk's commitment to colonizing the planet by 2050, for example, speaks volumes about our enduring fascination with the planet and emphasizes the value of asking the "who?" questions. ■

The reviewer is at the Department of English, University of Illinois, Urbana-Champaign, Urbana, IL 61801, USA. Email: rmarkley@illinois.edu

10.1126/science.adh7737

ECOLOGY

Voyage of the botanists

A reporter recounts the tale of a daring expedition that yielded vital insights about the Grand Canyon's flora



Brave the Wild River: The Untold Story of Two Women Who Mapped the Botany of the Grand Canyon
Melissa L. Sevigny
Norton, 2023. 304 pp.

By **Barbara J. King**

On 20 June 1938, a trio of small boats carrying six people set off from Green River, Utah, with the goal of running the Colorado River through the Grand Canyon. No previous Colorado River expedition, not even those led by famed explorer John Wesley Powell, excited national news coverage in quite the way this one did, for this river party included two women. The brainchild of University of Michigan botanist Elzada Clover, who invited doctoral student Lois Jotter to join, this expedition was scientific, focused on plants. No woman had done anything like this before.

In *Brave the Wild River*, science journalist Melissa Sevigny brings the voyage compellingly to life. Through close archival research, she traces the 43-day journey from Green River through the Grand Canyon to Lake Mead and the then-new Boulder (now Hoover) Dam. A page-turner in the adventure genre that also conveys rich detail about plant ecology of the US Southwest, sexism in science, and ethical issues in environmental science, the book is a marvelous and informative read.

At 41, Clover was a newly minted PhD and devotee of plant collecting and cataloging. It was her dream to map the plants of the Southwest. On a 1937 trip that took her through Mexican Hat, Utah, she met commercial river rafter Norman Nevills, who agreed to take on the expedition. Jotter, 23, was deep into her own doctoral research focused on cytogenetics, an emerging field focused on the relationship between chromosomes and cell behavior. With the addition of three crew members, the group set off down the Green River and into the mighty Colorado, then a river of wild frothing rapids that on more than one occasion put the expedition in danger.

Botany, in the 19th century, was the single branch of natural history considered suitable for women, observes Sevigny. "Botany's virtues were women's virtues: gentility, innocence, and modesty." Yet limitations were ever present. Women, for example, "were not permitted to collect any plant that couldn't be reached on a leisurely afternoon stroll."

Around the start of the 20th century, laboratory analysis of plants became the fashion, and the field became overwhelmingly male-dominated. Among academic botanists, Clover and Jotter were a rarity—and they hungered to create a plant collection so far unknown to Western science.



The party arrives at Lee's Ferry in Arizona.

Sevigny captures the women's excitement at amassing desert-adapted specimens and preparing them for shipment back to Michigan. They collected everything from four-wing saltbush and strawberry hedgehog cacti to ocotillo and Southern maidenhair fern. "In cataloging plants that thrived in extremes, they were adding to the general picture of evolution and adaptation," she writes. Non-native plants were found too. The tumbleweed, for example, had been introduced in bushels of flaxseed imported from Russia in the 1870s and had spread across 35,000 square miles in only 20 years.

Even as Clover and Jotter were making scientific history, they faced the full force of gender expectations and stereotypes. Each morning on the river, they cooked breakfast not just for themselves but for the entire crew. Frank Dodge, who had run the Colorado on earlier expeditions, referred to Clover as "a spinster" and Jotter as a "nice buxsom gal." Still, the women's work made a lasting contribution.

Although they did not use "ecosystem" language in describing their project, the pair nonetheless proceeded within a recognizably ecological framework. They worked out how plants dispersed throughout the desert by means of the movement of water and animals as diverse as mice and bighorn sheep. Their 1944 paper in the journal *American Midland Naturalist* listing >400 plant species has influenced "generations of botanists and ecologists in the Southwest."

In 1994, when Jotter was 83 years old, the US Geological Survey invited her to join a 12-day rafting trip aimed at tracking changes in Grand Canyon ecology. Game for the adventure, she encountered a river that flowed in a far steadier and more predictable way than it had in 1938.

It was not, of course, that the area had in earlier times remained free of human influence. Sevigny explains throughout the book that the canyonlands had long been home to Indigenous peoples, such as the Havasupai, who successfully hunted, gathered, and farmed the desert and then suffered at the hands of white settlers. This Colorado River was (and is) different and tamer. More dams had been erected, and the river had become a source of hydroelectric power, a water-delivery system for far-flung communities, and a "playground for tourists."

People's needs and preferences have taken fiercer priority over the health of this interrelated community of plants and animals, which have suffered because the river's natural rhythms have become so altered. From that final boat trip of Jotter's through to the current terrible drought in the Southwest, we see that it is time, as Sevigny puts it, "to imagine new kinds of future in which human desires don't subsume a river's needs." ■

10.1126/science.adf8933

The reviewer is professor emerita of anthropology at William & Mary, Williamsburg, VA 23185, USA. Email: bjking@wm.edu

LETTERS



Illegal mining has decimated ecosystems in Canaima National Park, Venezuela.

PHOTO: ZABELA STACHOWICZ

Edited by **Jennifer Sills**

Venezuela's harmful mining activities grow

In 2016, Venezuela's government designated 112,000 km^2 of tropical rainforest and Indigenous land as the Orinoco Mining Arc (OMA), a special economic zone for mining (*1*). This controversial decision, which aimed to tighten government oversight over mineral mining and trading in the region, resulted instead in promoting a complex network of intertwined legal and illegal mining activities throughout the country (*2*). To protect Venezuela's people, culture, biodiversity, and environment, the country must curtail mineral extraction.

Since the OMA was established, there has been over 60,000 ha of mining activity in the southern Orinoco region (*3*). In the Guiana Shield, one of the most biodiverse regions in the world, mining is responsible for more than 90% of total deforestation (*4*). The mines are subject to no environmental or sociocultural impact assessments (*3*). Moreover, designating the land for mining in the OMA violates the national sovereignty and natural and cultural heritage of 14 Indigenous Nations (*5*) as well as the rights of Venezuelan citizens to prior consultation.

The OMA's lack of oversight has allowed mining to proliferate in protected areas (*6*). Mining has led to the destruction of ecosystems in the Canaima and Caura National Parks, which overlap with the OMA. Yapacana National Park, which is located outside the OMA's borders, has been harmed indirectly (*7*). Mining activities are also taking place atop sacred tepuis (*8*), table-top mountains considered houses of gods by Indigenous peoples.

Mining in the OMA has also exacerbated violence and public health challenges. Military and armed syndicates have been fighting for control of various mining operations, further blurring the lines between legal and illegal activities (*9*). The mining workforce in the OMA includes child labor and modern slavery. Violence and crimes against women are commonplace (*2*). More than 80% of malaria cases in regions bordering the Brazilian Amazon originate in Venezuela in areas with high mining activity and increasing deforestation (*10*).

The negative impacts of mining on society and international security are overlooked in most long-term socioeconomic planning contexts, in part because of critical data gaps. In Venezuela, official data on mining extraction are lacking (*6*), and misleading global statistics hinder control measures and investigative processes.

To mitigate the social and ecological damage caused by mining in Venezuela, the international community should monitor Venezuela's mining activities (*11*) and, in the case of illegal activities, penalize those involved. The Venezuelan government should work with neighboring countries to eliminate the illicit market and require international organizations to reduce the purchase of illegally extracted gold. The government should also consult with civil society, especially Indigenous (*12*) and scientific communities, to assess the risks of mining activities and devise and enact mitigation and environmental restoration plans.

Izabela Stachowicz[1]*, Vilisa Morón Zambrano[2], Anthony J. Giordano[3], José R. Ferrer-Paris[4], Stefan Kreft[5]

[1]Department of Geobotany and Plant Ecology, University of Łódź, Łódź, Poland. [2]Sociedad Venezolana de Ecología, Caracas, Venezuela. [3]S.P.E.C.I.E.S. (The Society for the Preservation of Endangered Carnivores and their International Ecological Study), Ventura, CA, USA. [4]University of New South Wales, School of Biological, Earth, and Environmental Sciences, Sydney, Australia. [5]Society for Conservation Biology–Latin America and Caribbean Section, Washington, DC, USA.
*Corresponding author.
Email: stachowicz.izabela@gmail.com

REFERENCES AND NOTES

1. "Decreto N° 2.248, mediante el cual se crea la Zona de Desarrollo Estratégico Nacional 'Arco Minero del Orinoco'" (2016); https://www.notilogia.com/2016/02/gaceta-oficial-n-40-855-creacion-del-arco-minero-del-orinoco.html [in Spanish].
2. B. Ebus, "Digging into the Mining Arc," *Info Amazonia and Correo del Caroni* (2018); https://arcominero.infoamazonia.org/story.
3. "Amazonía bajo Presión 2020" (Red Amazónica de Información Socioambiental Georreferenciada, 2020); https://www.raisg.org/es/publicacion/amazonia-bajo-presion-2020/.
4. M. Kalamandeen *et al*., *J. Appl. Ecol*. **57**, 1730 (2020).
5. V. Morón-Zambrano, E. Isasi-Catalá, F. Cavada-Blanco, E. Agudo-Adriani, A. Blanco-Davila, Eds., *Una Mirada al Soberbio Sur del Orinoco: entendiendo las implicaciones del Arco Minero: Edición Especial de Explora* (Sociedad Venezolana de Ecología y Provita, 2018) [in Spanish].
6. "Gold flows from Venezuela: Supporting due diligence on the production and trade of gold" (OECD, 2021); https://mneguidelines.oecd.org/gold-flows-from-venezuela-supporting-due-diligence.htm.
7. M. R. Cabello, "Gold mining devastation beneath the eyes of Roraima Tepuy," *Correo del Caroni* (2020).
8. S. Schmidt, A. V. Herrero, J. K. Chen, "The gold-mining city that is destroying a sacred Venezuelan mountain," *The Washington Post* (2022).
9. UN Office of the High Commissioner for Human Rights, "Venezuela: UN releases report on criminal control of mining area and wider justice issues" (2020); https://www.ohchr.org/en/press-releases/2020/07/venezuela-un-releases-report-criminal-control-mining-area-and-wider-justice?LangID=E&NewsID=26096.
10. M. E. Grillet *et al*., *Science* **359**, 528 (2018).
11. C. V. Burelli, "Venezuela's ecocide needs international attention" (Center for Strategic and International Studies, 2020).
12. Convention on Biological Diversity, "Kunming-Montreal Global Biodiversity Framework: 2030 targets" (2022); https://www.cbd.int/gbf/targets/.

10.1126/science.adh4314

Renewable energy in China's abandoned mines

Maximizing the development of renewable energy such as wind and solar power is an effective way to achieve carbon neutrality (*1*). China has promised to triple its wind and solar capacity to more than 1.2 GW by 2030 (*2*), but the photovoltaic and fan equipment needed to meet this goal will require substantial land resources (*3*). Although the country is building massive wind and solar power bases in the western Gobi Desert, the desert region is far from areas of high power demand (*4*). China needs to find more land for renewable energy systems in the eastern part of the country (*5*). Repurposing abandoned mine lands could be the solution (*6*).

Scientific and governmental interest in land-constrained energy production is growing (*7*). Chevron Questa has built photovoltaic projects in an open-pit mine in New Mexico (*7*). Photovoltaic projects have also been initiated in the abandoned mines in Meuro and Schipkau, Germany (*8*).

China has almost 13,000 abandoned coal mines spread across the country (*9*). Approximately 23,000 km^2 of these lands, including subsidence area and abandoned land, are suitable for the construction of photovoltaic power plants (*10*). The topology of coal mines makes them particularly well matched to the needs of pumped-storage power stations—the most widespread and advanced method of storing electricity and adjusting voltages (*11*). Hydropower plants require underground space and water, both of which already exist in abandoned mines. Expanding development to the available lands could replace approximately 23% of China's coal-fired power plants (*10*) and improve the efficiency and reliability of distributed power generation systems (*11*, *12*).

Project plans should take the risks of abandoned mine lands into account. For example, some mine goafs (areas where minerals have been extracted) lack stability. Before construction begins, safety assessments should be conducted. If necessary, the grouting can be reinforced and the underground structure can be strengthened (*12*). Once stability is confirmed, most abandoned mines will have considerable potential for developing renewable energy in China.

Gang Lin[1,2], Yanan Zhao[1], Jingying Fu[1,2,3], Dong Jiang[1,2,3]*

[1]Institute of Geographic Science and Natural Resources Research, Chinese Academy of Sciences, Beijing, China. [2]College of Resources and Environment, University of Chinese Academy of Sciences, Beijing, China. [3]Key Laboratory of Coupling Processes and Effects of Natural Resource Elements, Ministry of Natural Resources, Beijing, China.
*Corresponding author. Email: jiangd@igsnrr.ac.cn

REFERENCES AND NOTES

1. H. Duan *et al*., *Science* **372**, 378 (2021).
2. The State Council of the People's Republic of China, "Carbon Peak Action Plan by 2030" (2021); http://www.amac.org.cn/businessservices_2025/ywfw_esg/esgzc/zczgsc/202112/P020211216597927285031.pdf [in Chinese].
3. S. Dunnett, R. A. Holland, G. Taylor, F. Eigenbrod, *Pro Natl. Acad. Sci. U.S.A.* **119**, e2104764119 (2022).
4. C. Wang *et al*., *Renew. Sustain. Energ. Rev*. **134**, 110337 (2020).
5. Y. Yang, S. Xia, *Science* **378**, 609 (2022).
6. B. Niblick, A. E. Landis, *Renew. Sustain. Energ. Rev*. **60**, 489 (2016).
7. United States Environmental Protection Agency, "RE-Powering America's land" (2023); https://www.epa.gov/re-powering.
8. List Solar, Largest solar power parks in Germany (2021); https://list.solar/plants/largest-plants/solar-plants-germany/.
9. X. Lyu *et al*., *Sci. Tot. Environ*. **822**, 153646 (2022).
10. S. M. Quan, "The potential of developing renewable energy on abandoned mine lands in China," thesis, University of Chinese Academy of Sciences, Beijing (2019) [in Chinese].
11. R. Gao, F. Wu, Q. Zou, J. Chen, *Energy* **243**, 123061 (2022).
12. J. Fan *et al*., *Appl. Energ*. **258**, 114007 (2020).

10.1126/science.adi1496

PAST AS PROLOGUE

Embracing questions after my father's murder

My father was stabbed and killed by a home intruder when I was 8 years old. I found him shortly before he died. That night, I could not conceive that a person could have hurt him, nor did it occur to me that the pool of blood he was lying in was his own. It plagued me for years that my last words to him were not "I love you" or "I will never forget you." Instead, standing over his fading body, I repeatedly asked him, "What happened?"



The author's father, shown here, encouraged curiosity.

For 30 years, I've grappled with my memory of that night and the feelings of guilt that arose because my last words to my father were a question. Now, as an independent scientist, I understand that I should not regret my instinct to ask questions; rather, searching for answers is the mark of a scientist. Throughout my life, I have continued to ask questions, striving to discover new information and explain the world around me.

As I make peace with the past, I have also realized that I have the strength to pursue even more information about that defining event. I recently submitted my first grant to study the epidemiology of fatal stab wounds—a substantial departure from my current line of research. I hope that this work will eventually lead to improved management and prevention of stab wounds, as well as more training for front line workers.

I love being a scientist, and it is now clear to me that my childhood experiences set me on this path even before that tragic night. My father always encouraged me to be curious. My early years were filled with exploration and the quest for knowledge. Today the memory of my father still pushes me to be inquisitive, skeptical, and passionate. I will continue to ask questions with pride.

Jacquelyn J. Cragg
Collaboration for Outcomes Research and Evaluation, Faculty of Pharmaceutical Sciences & International Collaboration on Repair Discoveries, University of British Columbia, Vancouver, BC, Canada. Email: jacquelyn.cragg@icord.org

10.1126/science.adh5157

Call for Submissions Past as Prologue is an occasional feature highlighting the role of family history in the life of scientists. What role did your family background play in your decision to pursue science, your field, or your career? Submit your story to www.submit2science.org.

TECHNICAL COMMENT ABSTRACTS

Comment on "Fairness considerations in global mitigation investments"

Gregor Semieniuk, Jayati Ghosh, Nancy Folbre

In an otherwise excellent analysis of fair regional shares of global mitigation investments, Pachauri *et al*. (Policy Forum, 9 December 2022, p. 1057) dramatically overestimate developing countries' "capability" to invest by estimating GDP using purchasing power parity exchange rates. Since internationally sourced investment goods must be paid for at market exchange rates, capability-based interregional finance flows should be vastly larger.

Full text: dx.doi.org/10.1126/science.adg5893

Response to Comment on "Fairness considerations in global mitigation investments"

Shonali Pachauri, Setu Pelz, Christoph Bertram, Narasimha D. Rao, Keywan Riahi

We welcome the analysis of Semieniuk *et al*. as an additional sensitivity to illustrate a more extreme distribution of regional contributions to climate mitigation investments that supports our main conclusion regarding the North-South divide in mitigation investment capabilities. In response to Semieniuk *et al*. we would like to first point out that, in defining the required global mitigation investments for the 2020 to 2030 period, our study relies on the estimates in the sixth assessment report of the Intergovernmental Panel on Climate Change (IPCC) WGIII. These are based on diverse sources and underlying models that to varying degrees reflect regional differences in technology costs and consider both purchasing power parity and market exchange rates. We use these IPCC estimates as a starting point and focus entirely on the question of how much of the needed regional investments, given different fairness considerations, should be financed from sources within a region.

Full text: dx.doi.org/10.1126/science.adh1463

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**



QUANTUM OPTICS

Entangling microwaves with optical photons

Several platforms are under development for quantum computation, simulation, and metrology applications, with each platform operating at different operational wavelengths for optimized performance. For practical technologies, the reality will likely be a hybrid of platforms that require quantum entanglement to be generated and shared across platforms with a large energy disparity. Sahu *et al*. introduce an electro-optical device that allows the generation of quantum entanglement between microwaves (the operational wavelengths of superconducting circuits) with optical photons (the operational wavelength of long-distance quantum communication). Bridging platforms with more than five orders of magnitude difference in energy scales and maintaining the fragile entanglement provides a route to efficiently linking up hybrid quantum systems. —ISO *Science*, adg3812, this issue p. 718

Artistic rendering of a device that produces quantum entanglement between microwaves and optical photons

ILLUSTRATION: ELI KRANTZ, KRANTZ NANOART

HEART REGENERATION

Cardiac redifferentiation and maturations

Unlike mammals, some animal species such as zebrafish can regenerate their heart after injury by using their surviving cardiomyocytes to produce new cells. It is unclear how these cells stop dividing and mature enough to contribute to heart function. Nguyen *et al*. used ex vivo imaging to visualize intracellular calcium dynamics in the regenerating zebrafish heart to study cardiomyocyte maturation. The formation of the cardiac dyad, a structure responsible for regulating intracellular calcium handling, played a critical role in determining whether cardiomyocytes proliferated or progressed through maturation. Furthermore, LRRC10, a component of the cardiac dyad, was implicated in the regulation of cardiomyocyte maturation in zebrafish, mouse, and humans. —SMH

Science, abo6718, this issue p. 758

CARBON SEQUESTRATION

Leaving forests alone

Harnessing the carbon-capturing potential of forests is a key component of plans to mitigate global climate change. Planting new forests is a common strategy, but this approach can have negative social and ecological impacts and substantial costs. Roebroek *et al*. instead investigated how ceasing management (e.g., wood harvesting or fire suppression) of forests would change their global carbon sequestration capacity. The authors assessed the differences between the biomass of similar forests with and without human activities and used machine learning to predict the additional biomass gain from removing human activities from global forests. Even if all management ceased (an extremely unlikely scenario), global forest carbon would only increase by about 15%. This work provides further evidence that changing forest management is not an alternative to cutting carbon emissions. —BEL

Science, add5878, this issue p. 749

CATALYSIS

Directing light-olefin synthesis

The production of light olefins from a mixture of hydrogen and carbon monoxide (syngas) over metal-oxide zeolite catalysts is complicated by competing side reactions. Jiao *et al*. show that the Brønsted acid sites created in the zeolite AlPO-18 by germanium substitution were highly active for carbon–carbon coupling of the intermediates to olefins but much weaker for the unwanted size reactions of hydrogenation and olefin oligomerization. A carbon monoxide conversion of 85%, along with a selectivity of for light olefins of 83%, led to an overall yield of almost 50%. —PDS

Science, adg2491, this issue p. 727

CALORICS

A cool bundle of tubes

Vapor compression cooling often relies on refrigerants that are greenhouse gases or have other issues with flammability and toxicity. Caloric cooling is a different strategy that instead relies on moving solids through a phase

transition. Qian *et al*. developed an elastocaloric cooling device that compresses fatigue-resistant bundles of nickel–titanium tubes to obtain an attractive cooling power and maximum temperature difference. The device is competitive relative to other caloric strategies and may be attractive for eventual commercialization. —BG

Science, adg7043, this issue p. 722

CIRCADIAN RHYTHM

Cardiometabolic sex differences under stress

Working night shifts increases the risk of cardiometabolic dysfunction. Studying mice on a high-fat diet and an altered circadian schedule similar to that of nighttime shiftwork, Anderson *et al*. found that female mice were more resilient to the cardiometabolic consequences of the regimen compared with male mice, which showed stress-related changes at both the tissue and microbiome levels. Sexual dimorphism in cardiometabolic dysfunction was replicated in a human cohort of shift workers, and this effect might in part explain why men are more prone to developing metabolic disease. —CAC

Sci. Transl. Med. (2023) 10.1126/scitranslmed.abo2022

BIOENGINEERING

Restoring naturalistic thermal sensation

Recent advances in prosthesis development for upper-limb amputees have shown that it is possible to restore touch sensation using different approaches. However, current technologies do not allow the restoration of thermal sensation. To provide upper-limb amputees with a more natural feeling, Iberite *et al*. developed a noninvasive wearable device able to restore thermal sensation by delivering thermal stimuli in sensitive areas of the residual limb. Using this approach, they were able to elicit thermal phantom sensations. The amputees were able to discern different thermal stimuli reliably and consistently. The restoration of thermal sensation, in combination with existing technologies, will contribute to providing a close-to-natural experience in prosthesis users. —MMa

Science, adf6121, this issue p. 731

MICROBIAL TOXINS

Junction dysfunction

The foodborne pathogen *Bacillus cereus* damages intestinal epithelial cells through various secreted toxins, such as the pore-forming toxin alveolysin. Sun *et al*. identified a strain of *B. cereus* that uses alveolysin to compromise the barrier function of the mouse intestinal epithelium and colonic epithelial cell monolayers. Alveolysin persistently disrupted cell-cell junctions by increasing the production of the microtubule-binding protein CFAP100. CFAP100 promoted microtubule polymerization and stability, resulting in disorganization of the microtubule network and junction impairment. —AMV

Sci. Signal. (2023) 10.1126/scisignal.ade8111



A new small, wearable device provides thermal sensation for amputees.

IN OTHER JOURNALS

Edited by **Caroline Ash** *and* **Jesse Smith**

ECONOMICS

Searching for jobs, finding gender gaps

Gender differences in how the first postcollege job hunt is navigated contribute to women earning less than men early in their careers. Drawing upon survey data from recent college graduates, Cortés *et al.* show that women accepted job offers earlier than men on average, and for lower pay. Sizable proportions of these gaps were explained by women being more risk averse and men being more overconfident in their earnings expectations, findings corroborated by laboratory experiments. Over time, men accepted jobs aligned with lower, more realistic earnings expectations, narrowing the earnings gender gap. —BW

Q. J. Econ. (2023) 10.1093/qje/qjad017



PHOTOSYNTHESIS

Engineering carbon fixation

The plant carbon dioxide (CO_2)–fixation enzyme Rubisco has low catalytic efficiency. However, some bacteria have mechanisms to internally concentrate CO_2 by surrounding Rubisco in a polyhedral protein complex called the carboxysome. Chen *et al.* introduced nine carboxysome genes from *Halothiobacillus neapolitanus* into tobacco chloroplasts, which do not normally have a CO_2-concentrating mechanism. The authors observed functional carboxysomes forming in tobacco with greater Rubisco levels than was found in previous attempts. When carboxysomes replaced native tobacco Rubisco, plants were able to grow, albeit in an elevated CO_2 environment. Future improvements, for example,

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

ANTHROPOLOGY

Kissing as a vector of disease transmission

There are two main types of kissing: friendly-parental and romantic-sexual. Although friendly-parental kissing appears to be ubiquitous among human societies past and present, romantic-sexual kissing is not a universal behavior. Additionally, the emergence of sexual-romantic kissing may have had a secondary effect on disease transmission, propagating the spread and evolution of orally transmitted pathogens. In a Perspective, Arbøll and Rasmussen draw from case studies in Mesopotamia of medical records, artwork, and ancient DNA analysis to posit that sexual-romantic kissing was probably common practice in ancient times and did not have a singular origin. The authors explain why such kissing is not likely to have contributed to changing patterns of disease transmission in ancient cultures. —GKA

Science, adf0512, this issue p. 688

CONSERVATION AND LAW

The use of science in rights-of-nature laws

Laws that establish legal rights for nature are being pursued in a growing number of countries to protect the environment. The success or failure of these rights-of-nature laws can depend in large part on how scientific concepts and expertise have been used to develop, interpret, and implement them. Epstein *et al.* reviewed key scientific aspects of rights-of-nature laws and the use of science in court decisions that have interpreted them. They examined the "right to evolve" to illustrate challenges in applying scientific concepts in rights-of-nature laws and identify some possible solutions. —BW

Science, adf4155, this issue p. 704

HUMAN GENOMICS

Making sense of genetic findings

Genome-wide association studies (GWASs) identify links between individual gene variants and various traits and diseases. Unfortunately, the findings from these studies cannot be used to determine whether the gene variants associated with a disease directly cause the condition or just happen to be located near biologically relevant genes or regulatory regions. Most of the variants identified through GWASs are located in noncoding regions of the genome, further increasing the difficulty of interpretation. A workflow developed by Morris *et al*. addresses this problem by using CRISPR-based editing to directly introduce variants of interest and then assessing their effects on gene expression in individual cells, thereby identifying their contributions to specific blood cell traits. —YN

Science, adh7699, this issue p. 705

CHEMICAL PHYSICS

Imaging the onset of symmetry breaking

The Jahn-Teller (JT) effect is a fundamental mechanism of geometric relaxation that leads to symmetry breaking in nonlinear molecular systems in degenerate electronic configurations. Despite being responsible for a variety of phenomena observed in many fields, direct imaging of this effect with high spectral resolution has remained a grand challenge. Using a combination of state-of-the-art experimental and theoretical time-resolved x-ray absorption techniques, Ridente *et al*. successfully reconstructed the ultrafast (few-femtosecond) molecular dynamics arising from the JT distortion in the prototypical example of the methane cation, as well as subsequent coherence and dissipation of energy to other internal degrees of freedom. In the future, the presented techniques could be used to study the JT distortion and other related dynamics in ultrashort time scales in more complex systems. —YS

Science, adg4421, this issue p. 763

BIOMATERIALS

A truer skin mimic

Our skin provides a protective layer for our bodies, but it also enables detailed sensory feedback and soft interactions with our surroundings. Wang *et al*. devised a prosthetic electronic skin that incorporates organic semiconductor transistors and has no rigid components, thus mimicking the mechanical aspects of real skin (see the Perspective by Sekitani). At the same time, it can sense external stimuli such as temperature and pressure and encode these stimuli into electrical pulses. The authors showed that the prosthetic skin could evoke neuronal firings at the motor cortex in a rat in vivo, which triggered toe twitching. —MSL

Science, ade0086, this issue p. 735; see also adf0262, p. 690

GLOBAL HYDROLOGY

Losing lake water

The amount of water stored in large lakes has decreased over the past three decades due to both human and climatic drivers. Yao *et al*. used satellite observations, climate models, and hydrologic models to show that more than 50% of both large natural lakes and reservoirs experienced volume loss over this time (see the Perspective by Cooley). Their findings underscore the importance of better water management to protect essential ecosystem services such as freshwater storage, food supply, waterbird habitat, cycling of pollutants and nutrients, and recreation. —HJS

Science, abo2812, this issue p. 743; see also adi0992, p. 693

DENDRITIC CELLS

Cholesterol triggers tolerogenic DCs

Dendritic cells (DCs) rapidly undergo immunogenic maturation in response to pathogen exposure, but the signals driving their homeostatic maturation into tolerogenic DCs that help to prevent autoimmunity remain unclear. Within the mouse spleen, Bosteels *et al*. identified engulfment of apoptotic cells as a key driver of the homeostatic maturation program among type 1 conventional DCs (cDC1s). Using single-cell transcriptomics, the authors found that pathways controlling cholesterol efflux, including the liver X receptor pathway, were induced during homeostatic cDC1 maturation. Cholesterol uptake through either apoptotic cells or unadjuvanted lipid nanoparticles promoted homeostatic DC maturation. Conversely, DCs lacking the key receptors initiated an immunogenic maturation program after engulfment of apoptotic cells, demonstrating that cholesterol-regulatory pathways control whether immunogenic or tolerogenic DC maturation occurs. —CO

Sci. Immunol. (2023) 10.1126/sciimmunol.add3955

ASYMMETRIC CATALYSIS

Screening for breadth

When chemists optimize catalysts, they typically work with a single substrate in a test reaction. Then, the catalyst producing the best outcome with that substrate is applied to a broader range of substrates. Rein *et al*. report a different approach: screening catalysts concurrently on a diverse pool of substrates with the aim of optimizing median enantioselectivity across the pool. Structural modification of a peptide-substituted aminoxyl radical framework using this method efficiently pinpointed a broadly versatile

catalyst for the oxidative desymmetrization of diols. —JSY

Science, adf6177, this issue p. 706

BEHAVIORAL ECOLOGY

False sense of security

In intact ecosystems, large carnivores have a regulating influence on smaller predators (mesopredators), limiting both their populations and distribution. In areas where humans are present, these smaller carnivores use developed areas to avoid larger predators. However, Prugh *et al*. found that their impression of the safety of these areas is a mistake, because the mortality rates of these species were more than three times higher than in the presence of large carnivores (see the Perspective by Darimont and Shukla). Such errors could both threaten these smaller species and influence the trophic structure of ecosystems. —SNV

Science, adf2472, this issue p. 754;
see also adh9166, p. 691

transition. Qian *et al*. developed an elastocaloric cooling device that compresses fatigue-resistant bundles of nickel–titanium tubes to obtain an attractive cooling power and maximum temperature difference. The device is competitive relative to other caloric strategies and may be attractive for eventual commercialization. —BG

Science, adg7043, this issue p. 722

CIRCADIAN RHYTHM

Cardiometabolic sex differences under stress

Working night shifts increases the risk of cardiometabolic dysfunction. Studying mice on a high-fat diet and an altered circadian schedule similar to that of nighttime shiftwork, Anderson *et al*. found that female mice were more resilient to the cardiometabolic consequences of the regimen compared with male mice, which showed stress-related changes at both the tissue and microbiome levels. Sexual dimorphism in cardiometabolic dysfunction was replicated in a human cohort of shift workers, and this effect might in part explain why men are more prone to developing metabolic disease. —CAC

Sci. Transl. Med. (2023) 10.1126/scitranslmed.abo2022

BIOENGINEERING

Restoring naturalistic thermal sensation

Recent advances in prosthesis development for upper-limb amputees have shown that it is possible to restore touch sensation using different approaches. However, current technologies do not allow the restoration of thermal sensation. To provide upper-limb amputees with a more natural feeling, Iberite *et al*. developed a noninvasive wearable device able to restore thermal sensation by delivering thermal stimuli in sensitive areas of the residual limb. Using this approach, they were able to elicit thermal phantom sensations. The amputees were able to discern different thermal stimuli reliably and consistently. The restoration of thermal sensation, in combination with existing technologies, will contribute to providing a close-to-natural experience in prosthesis users. —MMa

Science, adf6121, this issue p. 731

MICROBIAL TOXINS

Junction dysfunction

The foodborne pathogen *Bacillus cereus* damages intestinal epithelial cells through various secreted toxins, such as the pore-forming toxin alveolysin. Sun *et al*. identified a strain of *B. cereus* that uses alveolysin to compromise the barrier function of the mouse intestinal epithelium and colonic epithelial cell monolayers. Alveolysin persistently disrupted cell-cell junctions by increasing the production of the microtubule-binding protein CFAP100. CFAP100 promoted microtubule polymerization and stability, resulting in disorganization of the microtubule network and junction impairment. —AMV

Sci. Signal. (2023) 10.1126/scisignal.ade8111



A new small, wearable device provides thermal sensation for amputees.

IN OTHER JOURNALS

Edited by **Caroline Ash** *and* **Jesse Smith**

ECONOMICS

Searching for jobs, finding gender gaps

Gender differences in how the first postcollege job hunt is navigated contribute to women earning less than men early in their careers. Drawing upon survey data from recent college graduates, Cortés *et al.* show that women accepted job offers earlier than men on average, and for lower pay. Sizable proportions of these gaps were explained by women being more risk averse and men being more overconfident in their earnings expectations, findings corroborated by laboratory experiments. Over time, men accepted jobs aligned with lower, more realistic earnings expectations, narrowing the earnings gender gap. —BW

Q. J. Econ. (2023) 10.1093/qje/qjad017



PHOTOSYNTHESIS

Engineering carbon fixation

The plant carbon dioxide (CO_2)–fixation enzyme Rubisco has low catalytic efficiency. However, some bacteria have mechanisms to internally concentrate CO_2 by surrounding Rubisco in a polyhedral protein complex called the carboxysome. Chen *et al.* introduced nine carboxysome genes from *Halothiobacillus neapolitanus* into tobacco chloroplasts, which do not normally have a CO_2-concentrating mechanism. The authors observed functional carboxysomes forming in tobacco with greater Rubisco levels than was found in previous attempts. When carboxysomes replaced native tobacco Rubisco, plants were able to grow, albeit in an elevated CO_2 environment. Future improvements, for example,

introducing relevant bicarbonate transporters, may eventually allow carboxysomes to enhance carbon assimilation beyond endogenous levels. —MRS

Nat. Commun. (2023) 10.1038/s41467-023-37490-0

CELL BIOLOGY

How the NE is reassembled

The nuclear envelope (NE) that surrounds the nucleus is continuous with the endoplasmic reticulum (ER). Before mitosis, the NE disassembles and is reassembled immediately after cell division. Working in cultured human cell lines, Zhao *et al.* found that the NE can be assembled in a range of cell type–specific patterns. At one extreme, lateral sheet expansion involves the envelopment of peripheral chromatin by ER, followed by membrane extension from these sheets into chromatin within the spindle. At the other end of the spectrum, membrane infiltration involves the synchronous recruitment of smaller ER tubules or sheets to the entire chromatin surface by mitotic actin filaments. These NE assembly mechanisms result from cell type differences in ER morphology and the organization of mitotic actin. Thus, efficient postmitotic NE assembly requires different mechanisms tailored to cell type differences in mitotic endoplasmic reticulum morphology. —SMH

Dev. Cell (2023) 10.1016/j.devcel.2023.04.003

SIGNAL TRANSDUCTION

Closing up and repairing

Repair processes are inevitably required when organisms experience damage to their skin. Hypoxia occurs when blood flow is disrupted, and this can be sensed by mouse epithelial stem cells. Liu *et al.* found that in mice, hypoxia combines with basal signaling by interleukin-24 (IL-24), a homolog of interferon, to increase IL-24 autocrine and paracrine signals that regulate cell proliferation and metabolism. The metabolic effects increase the cellular production of lactate, which may extend the signaling effects to cells that do not express the IL-24 receptor. In this way, epithelial stem cells appear to coordinate their activation to close wounds with regulation of dermal and endothelial cells and fibroblasts that assist with the repair. —LBR

Cell (2023) 10.1016/j.cell.2023.03.031

MACHINE LEARNING

Tunneling processes in tropolone

Machine learning methods are gradually becoming the preferred approach for generating rapid and accurate full-dimensional potential energy surfaces (PES), a key component of molecular dynamics simulations in physics, chemistry, and materials science. Nandi *et al.* report a Δ-machine learned PES for tropolone, a 15-atom cyclic molecule, which improved the accuracy of a preexisting low-level PES to near CCSD(T) quality while maintaining its speed. The resulting PES was used in ring-polymer instanton calculations of the tunneling splittings of tropolone and isotopomers, which demonstrated excellent agreement with the experimental data and offered some useful and unexpected mechanistic insights into the process. The proposed combined methodology makes theoretical calculations of tunneling splittings robust and scalable to more complex chemical reactions in the future. —YS

J. Am. Chem. Soc. (2023) 10.1021/jacs.3c00769

GLOBAL WARMING

Seeing better from higher up

The pattern of tropospheric cooling accompanied by lower stratospheric cooling has long been recognized as a fingerprint of human disruption of the climate. Santer *et al.* have improved this approach by including data from the mid-to-upper stratosphere at altitudes of 25 to 50 kilometers. Extending the technique in this way greatly improved its sensitivity and showed definitively that satellite-measured trends in the thermal structure of the atmosphere must be anthropogenic, not natural as some have argued. —HJS

Proc. Natl. Acad. Sci. U. S. A. (2023) 10.1073/pnas.2300758120



PHOTO: ROBERT HENNO/ALAMY STOCK PHOTO

MICROBIOLOGY

Microbes and microhabitats

Microorganisms live in highly dynamic and heterogenous habitats. In the soil, microheterogeneity facilitates microbial density and diversity and, importantly, substrate preservation, which can scale up to influence carbon storage capacity. Arellano-Caicedo *et al.* developed microengineered arenas called fractal chips for experimental studies on the functional effects of the physical complexity of microbial habitats. The authors found that greater structural complexity limited fungal hyphal growth but permitted bacterial growth within the pore space. Further, when bacteria were present, the substrate was more fully degraded, possibly because enzymatic activity was concentrated and efficient in the small pore sizes deeper in the fractal mazes. However, far within the complex nooks and crannies of the fractal chips, where fungi and bacteria failed to penetrate, substrate turnover was slowest, which may allow long-term organic matter storage.

—CA *Curr. Biol.* (2023) 10.1016/j.cub.2023.02.064

Soil bacteria and fungi, such as *Coprinopsis cinerea,* whose fruiting bodies are shown here, live in different pore spaces.

REVIEW SUMMARY

CONSERVATION AND LAW

Science and the legal rights of nature

Yaffa Epstein*, Aaron M. Ellison, Hugo Echeverría, Jessica K. Abbott

BACKGROUND: Over the past 15 years, legal rights for nature have gone from a fanciful idea to legal fact in a growing number of countries. To date, these laws have had the greatest legal impact in the Global South, where Indigenous activism, political conflicts, and lack of effective environmental laws have provided opportunity and incentive for legal experimentation, but they have also been enacted at the local level in North America. In late 2022, Spain enacted the first rights-of-nature law in Europe, establishing the legal personhood of the Mar Menor lagoon.

Rights-of-nature laws have reached a critical point at which they may either be normalized or marginalized. They have captured the public imagination, leading to growing advocacy for, and enactment of, these laws. Some of these laws have succeeded in protecting the environment, often with the aid of engagement from scientists who have helped to interpret and implement them. Others have failed, often not because of the concept that nature can have rights but due to lack of clarity—scientific or otherwise—about how the law should be applied. The engagement of scientists with these laws as they are enacted, implemented, and enforced has been a key factor in judges' ability to apply them.

ADVANCES: Rights-of-nature laws have been enacted in Ecuador, Bolivia, New Zealand, Uganda, Panama, Spain, and US and Canadian localities, among others. Courts in other jurisdictions including Colombia, Bangladesh, and India have also recognized rights of nature based on legal principles within those jurisdictions. These laws and court decisions necessarily reference scientific concepts in defining new right-holding entities and new rights for these entities. While some of these laws are so new that they have yet to be applied by courts, there have been dozens of court decisions in the past 5 years interpreting rights of nature laws. Ecuador is a prominent example: Judicial application of nature's constitutional rights has resulted in court decisions delineating rights for species, ecosystems, waterways, and even individual animals. Scientific concepts pertaining to rights of nature laws analyzed by courts include life cycles, evolutionary processes, ecological flows, and species, among many others. The level of judicial engagement with natural science in making these analyses has varied. While the Ecuador Constitutional Court has actively worked to involve natural scientists in its legal decision-making about the rights of natural entities, US courts have thus far declined to engage with scientific aspects of rights-of-nature laws, holding in one case that a natural entity's rights to "exist, flourish, and naturally evolve" were unconstitutionally vague. Spain's Mar Menor legal personhood law is being challenged in the Spanish Constitutional Court on similar grounds.

OUTLOOK: When courts have difficulty interpreting scientific concepts used in law, the law's ability to reach legislative goals is hindered. We identify areas where future research is needed. Because legal decision-makers alone may [illegible] have the expertise to understand what nature's legal rights entail, interdisciplinary research is urgently needed if these laws are to be meaningful. Legal and natural scientists should examine legal-scientific concepts such as the right to evolutionary processes in particular legal contexts, to clarify what type of scientific input is needed to protect this right for different types of natural entities. By contributing to interdisciplinary analyses of rights-of-nature laws before disputes arise, scientists can help contribute to the effectiveness of these laws. The availability of credible scholarly analysis of legal-scientific terms used in law would make these rights more tangible and accessible to the judges whose role it is to apply them. Although scientific uncertainty often cannot be eliminated, it's reduction in turn reduces legal uncertainty and thus helps meet the objection that rights-of-nature laws are too vague to be applied. Another type of interdisciplinary scholarship that would assist the functioning of rights-of-nature laws would be the examination of the duties of nature. Although some rights-of-nature laws grant rights for nature without corresponding duties, others equate nature to a legal person with both rights and duties. Uncertainty over liabilities and duties of nature has been an impediment to implementing some rights-of-nature laws. Scientists can help legal systems comprehend nature's potential legal obligations (e.g., "ecosystem services"), and what environmental protection measures may also be legally required to ensure natural entities can continue to fulfill these obligations. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: yaffa.epstein@jur.uu.se
Cite this article as Y. Epstein *et al.*, *Science* **380**, eadf4155 (2023). DOI: 10.1126/science.adf4155

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.adf4155



Mangroves in Ecuador, such as these at Black Turtle Cove on Santa Cruz, Galapagos, have legal rights that have been recognized in court.

REVIEW

CONSERVATION AND LAW

Science and the legal rights of nature

Yaffa Epstein[1,2]*, Aaron M. Ellison[3,4], Hugo Echeverría[5], Jessica K. Abbott[6]

We review the use of science by lawmakers and courts in implementing or rejecting legal rights for nature in Ecuador, India, the United States, and other jurisdictions where some type of rights of nature have been recognized in the legal system. We then use the "right to evolve" to exemplify how interdisciplinary work can (i) help courts effectively define what this right might entail; (ii) inform how it might be applied in different circumstances; and (iii) provide a template for how scientists and legal scholars can generate the interdisciplinary scholarship necessary to understand and implement the growing body of rights-of-nature laws, and environmental law more generally. We conclude by pointing to what further research is needed to understand and effectively implement the growing body of rights-of-nature laws.

Despite an ever-growing body of environmental laws, biodiversity loss and other catastrophic environmental damage continue to occur worldwide (*1*). Strengthening environmental laws and improving their enforcement is considered one of the keys to reversing this trend (*1*). Rights-of-nature laws are one instrument used in an increasing number of countries with a goal of improving legal systems' ability to protect the environment (*1*). These laws, which establish legal rights for nature or elements of it (e.g., species, ecosystems), have reached a critical point at which they may be either normalized or marginalized. They have engaged the public imagination, leading to growing advocacy for, and enactment of, these laws (Table 1) (*2*). Some of these laws have been used successfully to protect the environment, often with the aid of expertise from scientists who have helped interpret and implement them. Others have failed, often not because of the failure of the concept that nature can have rights but rather because of lack of clarity—scientific or otherwise—about how the law should be applied. For these laws to work as intended in the legal systems in which they reside, it is necessary for life and earth scientists, along with social and political scientists and others, to engage with these laws as they are enacted, implemented, and enforced (Fig. 1). Without scientific input at the various stages of environmental law making, such laws are more likely to fail to be effective, and if they fail too often rights-of-nature laws may cease to be enacted as politicians and civil society will think they are unworkable. With engagement from scientists however, as recent court decisions in Ecuador demonstrate, rights-of-nature laws are more likely to be upheld by courts and effectively implemented, which could in turn lead to the continued growth in influence and importance of these laws in protecting the natural world.

As scholars have examined in a variety of contexts, scientific information is often necessary to give effect to laws (*3*, *4*, *5*, *6*). This is true in all areas of law, such as tort law, in which scientific expertise is often needed to determine risk or causation (*7*), but it is especially true in environmental law (*3*). Scientific concepts which courts have had to interpret in recent years include, for example, genetically modified organism (*8*), point-source pollution (*9*), and climate change (*10*). These terms describe or use scientific ideas but when they are used in the law they also become legal terms (*4*). Differing interpretations of these terms and concepts in different disciplines not only affect specific decisions but also make it difficult for various parties to agree whether a given action complies with the law. How judges or other decision-makers understand, misunderstand, or ignore scientific concepts affects the outcomes of cases and the ability of the law to achieve its legislative goals (*11*).

The need to grapple with science and scientific uncertainty is a persistent conundrum in environmental law, and rights-of-nature laws present additional challenges. First, they use science in new contexts; it may be impossible to understand who the new rights holders are or what rights they have at all without contributions from scientists (*12*). Second, because these types of law seem unfamiliar to judges and decision-makers, they may be hesitant as to how and when to use science to apply them, even when the laws use similar terms or protect similar entities or processes as other environmental laws. The risks of decision-makers failing to properly apply scientific aspects of the new laws are therefore higher, as is the corresponding risk that the laws will be ineffective.

Clarification is needed regarding our use of "science," a term which itself does not have a universally agreed-upon definition (*13*). By



Fig. 1. Stages in the life cycle of a law where scientists' input can help rights-of-nature laws work.

[1]Uppsala University, Faculty of Law, Uppsala, Sweden.
[2]Swedish Collegium for Advanced Study, Uppsala, Sweden.
[3]Harvard University Herbaria, Cambridge, MA, USA.
[4]Sound Solutions for Sustainable Science, Boston, MA, USA.
[5]Universidad Hemisferios, Faculty of Law, Quito, Ecuador.
[6]Lund University, Department of Biology, Lund, Sweden.
*Corresponding author. Email: yaffa.epstein@jur.uu.se

science, we mean the organization of knowledge obtained through observations and experiments that leads to testable and refutable predictions about the causes of observable phenomena, and information that is produced by experts within disciplines that utilize these processes (*4*). Although this type of science is often referred to as "Western science", its origins in, for example, China, India, Egypt, and Mesopotamia transcend hemispheres and contemporary geographic boundaries (*14*). We acknowledge that other systems of knowledge, such as Indigenous Traditional Ecological Knowledge (ITEK) have and continue to make important contributions to the creation and interpretation of rights-of-nature laws, especially where rights of nature intersect with the rights and sovereignty of Indigenous Peoples (*15, 16, 17, 18, 19, 20*). "Western science" (hereafter referred to as "science") is frequently and necessarily applied to current legal formulations and applications of all types of rights-of-nature laws, however, particularly those that explicitly specify scientific terms and processes or create, for example, scientific advisory boards. There has been very little scholarly consideration of how science is used in formulating and applying rights-of-nature laws; we delimit our Review to addressing this gap in the literature.

Here, we review central scientific aspects of rights-of-nature laws and litigation. In the next section, we review some of the scientific concepts embedded within the laws themselves. We then examine the use of science in court decisions that have interpreted rights-of-nature laws. Finally, we use the example of the "right to evolve" to discuss some difficulties in applying scientific concepts in rights-of-nature laws and identify some possible solutions to these difficulties. We conclude with a call for further interdisciplinary collaborations between legal scholars and scientists to help courts and others understand legal "rights of nature" and their relationship with other environmental laws, so that all such laws can have their intended impact of protecting the environment.

Scientific concepts in rights-of-nature laws

Over the past 15 years, legal rights for nature have gone from a fanciful idea to a legal fact in at least 20 countries (*2*, *21*). To date, these laws have had the greatest legal impact in the Global South, where Indigenous activism, political conflicts, and lack of effective environmental laws provided opportunity and incentive for legal experimentation, but they have also attracted considerable attention in North America and Europe (*19*). Rights-of-nature laws take a variety of forms and grant rights to different types of rights holders, from the metaphysical "Mother Earth" to certain types of natural features such as rivers or ecosystems, to specifically delineated ecosystems (*22*). The types of rights recognized or afforded to natural entities also vary (*18*). Some laws assign natural entities legal personhood—a legal concept that accords entities the ability to have their rights and duties enforced by courts—similar to a corporation. Other laws enumerate substantive rights for natural entities, including property rights, the right to life, diversity of life, clean air, evolutionary capacity, protection, or restoration. Some of these rights, such as the right to property and life, resemble human rights whereas others such as the right to evolutionary capacity are unique to nonhuman natural entities (although humans also evolve, as of yet there is no recognized human right to evolve). In this section, we identify scientific concepts inherent in these rights-of-nature laws in a non-exhaustive set of examples and discuss the potential role and limitations of scientific input in implementing natures' rights.

Ecuador

Article 71 of Ecuador's constitution sets out rights of nature, or Pachamama, to "integral respect for its existence and for the maintenance and regeneration of its life cycles, structure, functions, and evolutionary processes," and the right to be restored. Further, the state has an obligation to take measures to prevent the "extinction of species, the destruction of ecosystems, and the permanent alteration of natural cycles" (*23*).

Pachamama is an Andean goddess associated with fertility and harvest and is used in the law to represent nature as a living whole

Table 1. Examples of rights-of-nature laws.

Ecuador Constitution Art. 71 (2008)	"Nature, or Pachamama, where life is reproduced and occurs, has the right to integral respect for its existence and for the maintenance and regeneration of its life cycles, structure, functions and evolutionary processes."
Bolivia Law 071 on the Rights of Mother Earth (2010)	Mother Earth has the right to life, which includes the "right to maintain the integrity of living systems and natural processes that sustain them, and capacities and conditions for regeneration," as well as rights to diversity of life, equilibrium, and restoration, among others.
Rights of Manoomin (Wild Rice), White Earth Band of Ojibwe (2018)	"Manoomin, or wild rice, within the White Earth Reservation possesses inherent rights to exist, flourish, regenerate, and evolve, as well as inherent rights to restoration, recovery, and preservation."
Uganda National Environment Act Art. 4 (2019)	"Nature has the right to exist, persist, maintain and regenerate its vital cycles, structure, functions and its processes in evolution."
Lake Erie Bill of Rights, Toledo, OH, USA (2019) (overturned by court decision)	Lake Erie has the "right to exist, flourish, and naturally evolve."
Recognition of the legal personality and rights of the Magpie River-Muteshekau-shipu Res. 025-21, Minganie Regional County Municipality, Quebec, Canada (2021)	The Magpie River has the fundamental rights, among others "to evolve naturally, to be preserved and to be protected"
Panama Law 287 recognizing the rights of nature and obligations of the state related to these rights (2022)	Nature has the right to exist, persist, and regenerate its life cycles; to timely and effective restoration; and to the preservation of its water cycles.
Rights of Mar Menor, Spain (2022)	The Mar Menor lagoon and its basin are recognized as having the rights to protection, conservation, maintenance and, if necessary, restoration, as well as the rights to exist as an ecosystem and to evolve naturally.

although, as argued by scholars, this legal Pachamama does not clearly correspond to Indigenous philosophies (19). The constitution's preamble also equates Pachamama with "Mother Earth" and states that we are a part of nature or Pachamama—i.e., humans are also part of this rights-bearing entity. Although it is unclear from the text how exactly this entity should be defined, it apparently contains elements that cannot be given content by or defined explicitly by scientists alone. For example, nature has the right to "respect for its existence." "Respect" may be defined sociologically whereas the meaning of "existence" can be debated by philosophers, scientists, and others. Other rights appear more amenable to scientific input or understanding including rights to life cycles, ecological functions, evolutionary processes, restoration, and prevention of species' extinctions.

New Zealand

In contrast to Ecuador, New Zealand does not recognize the rights of nature as a collective whole. Rather, New Zealand recognizes the rights of specific natural entities that are significant to particular Iwi (tribes). These entities are defined in both geographical and metaphysical terms (24) and have been granted legal personhood and property rights (18). The rights-bearing entity Te Awa Tupua, for example, is defined in a 2017 law as "an indivisible and living whole, comprising the Whanganui River from the mountains to the sea, incorporating all its physical and metaphysical elements" (25). As with Pachamama, science is needed to understand the physical and biological aspects of this new legal person but cannot solely define it.

Compared with many other rights-of-nature laws that have nature protection as a central goal, these acts are primarily aimed at protecting the rights of Indigenous Peoples and incorporating their conceptions of natural entities into the national legal system (26). For example, in vesting property rights in the Whanganui riverbed, the New Zealand government acknowledged that it did not own the natural entity itself. It further ceded decision-making authority to a body appointed to act in the interests of the river, comprised of representatives of the Indigenous Whanganui Iwi and the New Zealand government. ITEK may be of greater importance relative to Western science in applying the New Zealand rights-of-nature laws than some others, although defining the boundaries of the recognized legal persons may require reference to information derived from both ITEK and scientific epistemology (17) (27).

Panama

Panama's 2022 law on rights of nature sets out both nature's rights and the obligations of the state to protect them (28). The law defines "nature" as a "collective entity, indivisible and self-regulated, shaped by its elements, biodiversity, and interrelated ecosystems." No metaphysical concepts are referred to, but both scientific input and other expertise, such as ITEK, may contribute to defining this rights holder. Nature's rights are referred to as "fundamental" and thus comparable to human rights.

The law sets out a rather comprehensive set of rights, many of which contain scientific concepts. Some of these rights adhere to nature as a collective whole, whereas others "extend to all living beings, elements, and ecosystems of which [nature] is composed." Similar to several earlier rights-of-nature laws, these rights include "the right to exist, persist, and regenerate their vital cycles," the right to be restored, the right to air quality, the right to maintain its biodiversity, and "the right to regenerate its vital cycles … in such a way that it will not lead to an imbalance of the development and integral maintenance of its natural cycles nor evolution of ecosystems." Although scientific information is clearly needed to understand these rights, their formulation is problematic from an ecological perspective in that there is no guarantee that natural evolution would always act to maintain ecological balance. But given that the law requires this balance scientists may be called upon to illuminate how to maintain it.

Spain

In 2022, Spain enacted the first rights-of-nature law in Europe (29). This law recognizes the legal personality of the Mar Menor lagoon and its basin, including the "entire lagoon marine ecosystem," whose size, location, and specific aquifers are defined with geographical precision and in scientifically recognizable terms (30). The preamble to the law cites both ecological and humanitarian reasons for assigning legal rights to the lagoon, but, unlike the laws in Ecuador, New Zealand, and others, the definition is stated without reference to metaphysical concepts.

The rights accorded to Mar Menor include both legal personhood and substantive rights. The latter include the rights to protection, conservation, maintenance, and restoration, and the rights to exist as an ecosystem and to evolve naturally. What these rights entail are further legally defined. For example, the "right to exist and naturally evolve" is further clarified: "The Mar Menor is governed by a natural order or ecological law that makes it possible for it to exist as a lagoon ecosystem and as a terrestrial ecosystem in its basin. The right to exist means respecting this ecological law, to ensure the balance and regulation capacity of the ecosystem in the face of the imbalance caused by anthropic pressures coming mainly from the watershed." Although this gives some indication how the right to exist and evolve are to be applied, it also invites further scientific input in applying "ecological law," though what is meant by this term can likely not be defined by scientists alone. The Mar Menor law establishes several bodies to help protect the rights of Mar Menor, including scientific, representative, and monitoring committees. The scientific committee is tasked with advising the other committees, particularly about the ecological status of and threats to the ecosystem, as well as appropriate conservation measures. This law thus explicitly creates a pathway for scientific knowledge to interact with political and other expertise in protecting nature's rights.

USA

There are currently no state or national laws in the USA giving rights to nature, but dozens of rights-of-nature laws have been adopted at local and tribal levels. Below are two examples.

Santa Monica

Santa Monica's 2013 Sustainability Rights Ordinance recognizes the rights of "natural communities and ecosystems" to "exist and flourish" (31). The terms used to describe types of rights holders—natural communities and ecosystems—are used broadly by ecologists and defined further in the law as "groundwater aquifers, atmospheric systems, marine waters, and native species within the boundaries of the City." Ecologists, hydrologists, and geologists might help identify which natural communities and ecosystems meet the legal definition required by this ordinance.

White Earth nation

Several indigenous nations have enacted laws or resolutions pertaining to rights of nature. In 2018, for example, the White Earth Band of Ojibwe, located in Minnesota, recognized the rights of Manoomin, or wild rice, to exist, flourish, regenerate, and evolve, and its rights to restoration, recovery, and preservation (32). Ecologists and botanists could help identify, for example, the conditions necessary for wild rice to evolve.

A few local rights-of-nature laws have been overturned by courts for violating state laws or the US Constitution. However, the reason for most of these negative rulings was that many of these laws, which were drafted with the aid of a nongovernmental organization, were intended as "civil disobedience" (33). They frequently purported to supersede state law and strip corporations of legal personhood (33) and thus were in conflict with the US or state legal systems. Santa Monica's 2013 ordinance, however, avoided these conflicts and continues to build its environmental laws and policies based on these rights (34).

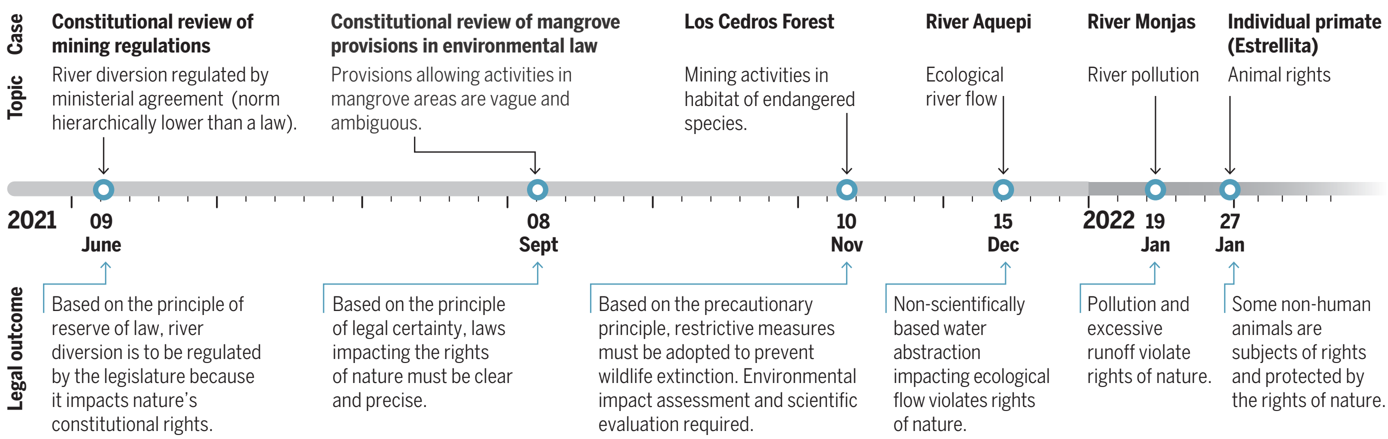


Fig. 2. Rights of Nature Cases in the Constitutional Court of Ecuador, 2021–2022.

Use of science in adjudicating the rights of nature

In this section, we discuss how scientific questions accompanying the implementation or enforcement of rights of nature have been dealt with by courts. Although rights for natural entities often include religious, spiritual, and social elements that may make it impossible to reduce them "to a host of ecological criteria" (*35*), it is necessary to also consider scientific criteria to interpret rights of nature. Even for rights that are analogous to recognized human rights such as the right to life, science may be needed to clarify similarly formulated rights of nature.

It is important to examine not only laws themselves but also how they are interpreted by courts. That is because laws alone, no matter how carefully drafted, cannot protect the environment unless they are also applied and enforced. It is often unclear what the impact of a law will be until it is tested in court. Although many rights of nature laws have not yet been litigated there have been dozens of relevant court decisions in the past 5 years, particularly in Ecuador, Bolivia, and Colombia. We review the use of science in several important examples of these decisions.

A judicial push to utilize scientific knowledge in Ecuador

As noted above, Ecuador was among the first countries to legally recognize rights for nature. For more than a decade after this recognition however, few cases on rights of nature were decided by courts, and it was unclear to what extent these constitutional provisions would affect the legal system or the environment (*2*). However, in 2019, a new Constitutional Court took office. In 2021 and 2022, Ecuador's Constitutional Court made six important decisions on how rights of nature should be applied (Fig. 2). The constitution protects the rights of nature as a whole to its life cycles, structure, functions and evolutionary processes but in many of these cases the Constitutional Court also recognized the rights of specific natural entities in order to identify the harms suffered as well as potential remedies (*36*). Scientific considerations and testimony have been essential to understanding and protecting the rights of nature and its interdependent elements.

Mining regulations case

One of the first rights-of-nature cases decided by the Constitutional Court was the June 2021 Mining Regulations case (*37*). This case considered whether administrative regulations for modifying or diverting watercourses to support mining activities violated the rights of nature by risking permanent damage to the seasonal rhythms and hydrologic regimes of the rivers, which could cause changes to the chemical composition of the water or other changes that could affect species and ecosystems (paragraphs 11 and 12). It was also alleged that such diversions would risk the ecological flows of the river, which were also protected under the 2008 Constitution.

In its decision, the court extensively analyzed natural science texts to understand ecological flows and adopted the following definition: "the quantity, quality, and timing of water flow needed to sustain ecosystems and the services they provide to humans" (paragraphs 56 to 61). Because diverting watercourses could affect ecological flows and thus constitutional rights, decisions to do so could not be made through simple administrative procedures, the court ruled, but rather through a decision-making process that took full account of the constitutional rights of nature. The court further required that future authorizations to divert watercourses must be based on independent technical studies that could ensure that the proposed diversion would not lead to the extinction of species, the destruction of ecosystems, or the permanent alteration of natural cycles (paragraph 80). This decision made clear that hydrological expertise would be needed to determine when modifications to a watercourse violate the rights of nature and other constitutional rights.

Los Cedros case

Los Cedros is a cloud forest of 6400 hectares. The bioregion of the Andes and the bioregion of Choco converge in Los Cedros, making it a high biodiversity area. It is the habitat of endangered flora and fauna including the critically endangered jaguar and spider monkey. In 2017, the State of Ecuador authorized exploration for industrial mining within the forest. This was challenged by the head of the local municipality. Although Los Cedros is designated as protected forest, mining was allowed in this type of designated area. The Constitutional Court's decision in the November 2021 Los Cedros case (*38*) extensively developed binding rights-of-nature jurisprudence and indicated to lower courts how future rights-of-nature cases should be analyzed and decided. The court expressed "much concern" that the rights of nature had not yet been given sufficient consideration in the legal system (paragraph 34). It saw its task in this case as deciding not whether certain mining permits were illegal in light of conservation law (on which the lower courts had focused), but rather whether the mining permits violated nature's rights to have its existence fully respected, or its right to maintain and regenerate its cycles, structure, functions, and evolutionary processes (paragraph 25). In adjudicating nature's rights, the court considered that it also needed to "analyze the rights of existence held by the animal and plant species of Los Cedros, as well as the right of this ecosystem to maintain its cycles, structure, functions and evolutionary process" (paragraph 26). In other words, the court considered that not only "nature" as a whole or

Pachamama had constitutional rights but that these rights extended to animal and plant species and the Los Cedros ecosystem.

This case made clear the importance of scientific evidence to adjudicating rights of nature. The court not only heard testimony from dozens of scientists and scholars but also "requested that national or international academic institutions that had performed scientific research on the Los Cedros Protected Forest…submit any information they considered relevant for the resolution of the [case]" (paragraph 4). The court began its analysis by carefully describing the Los Cedros ecosystem in "ecological terms" (paragraph 12). It then noted that "the characteristics of this ecosystem and the species that comprise it will be discussed further based on verified scientific information," and that this information would be considered "alongside the rights alleged to have been violated" (paragraph 13).

The court also noted that the constitutional rights of nature are "an intellectual convergence of the knowledge of indigenous peoples and modern Western science…and [in also referring to Mother Earth] recalls the essential relationship between human beings and nature" (paragraph 29). The court thus saw no contradiction in applying Western science to interpreting Pachamama's rights.

In one example of the court's incorporation of scientific concepts into legal analysis, the court considered that nature's right to "regeneration of its vital cycles, structure, functions and evolutionary processes" required understanding "the ecological tolerance principle, which holds that natural systems can only function adaptively within an environment whose basic characteristics have not been altered beyond what is optimal for that system…For each particular characteristic of the environment (amount of rain, humidity, solar radiation, etc.) there are limits beyond which organisms can no longer grow, reproduce, and ultimately survive. Thus, when the ecological tolerance level is exceeded it is impossible to exercise the right to reproduce life cycles" (paragraph 44). The court extensively referred to scientific sources to help it understand the scientific concepts necessary to interpret nature's constitutional rights.

On the necessity to protect species, the court applied the biological species concept (*39*) in its statement that "[t]he biotic component of the ecosystem is the species that constitute it. A species is defined as the set of organisms capable of interbreeding and producing fertile offspring…Often, over geological time, individuals that separate from the original population and become isolated from the rest may reach a sufficient degree of differentiation to become a new species" (paragraph 46). The court then argued that nature's right to existence meant protecting it from human activities that would lead to species extinction, which the court considered equivalent to genocide (paragraph 68).

Importantly, the court held that the precautionary principle, a principle which generally applies to threats to human health or the environment and which has been widely recognized in international agreements such as the 1992 Rio Declaration on the Environment and Human Development, must be applied when there is a potential risk of serious and irreversible damage to the rights of nature. This means, according to the court, that potentially damaging activities cannot be carried out unless there is scientific certainty that the rights of nature will not be irreparably harmed (paragraph 62). Therefore, scientific input will always be needed in determining whether human activities violate the rights of nature.

The integration of scientific, social, and spiritual data in India

No country-wide rights-of-nature laws have been enacted in India, but the highest courts in several states have recognized the rights of nature in court decisions. Similar to courts in several jurisdictions in other countries including Colombia and Bangladesh, these Indian courts considered that principles of law within their legal systems enabled or required them to recognize new legal persons.

Uttarakhand High Court

In its 2017 ruling, Mohd Salim *v.* State of Uttarakhand and others. The High Court of Uttarakhand at Nainital declared the Ganges and Yamuna Rivers and "all their tributaries, streams, every natural water flowing with flow continuously or intermittently of these rivers" to have "the status of a legal person with all corresponding rights, duties and liabilities of a living person in order to preserve and conserve" the rivers. Although the entities recognized were geographical features—waterways—the court, in its decision, emphasized the spiritual aspects of the two rivers: they are "sacred and revered" by Hindus. The state of Uttarakhand and a Ganges conservation organization were declared "in loco parentis" for the rivers, "as the human face to protect, conserve and preserve" the rivers and their tributaries. The court followed up ten days later with an additional decision, Lalit Miglani *v.* State of Uttarakhand and others, granting the same type of rights to the glaciers that feed the two rivers (*40*). This wide-ranging decision contained extensive literary descriptions of spiritual and cultural dimensions of nature in India. It also contained extensive quotations from scientific papers describing Himalayan ecosystems and biodiversity (pages 11 to 17).

The decision discussed the interdependence of natural entities in quasiecological terms. For example, "Trees are the buffer zone necessary to protect the glaciers from direct and indirect heat. One tree sustains life of thousand of insects. Birds chirp and make their nests on the trees. Trees are mini-reservoirs and have a capacity to store the water. The water stored by the trees is released slowly. The Oak tree preserves about 75,000 gallons of pure water. Plucking of one leaf, grass blade also damages the environment universally." The court also used complementary spiritual and scientific justifications for the rights granted: "Rivers and Lakes have intrinsic rights not to be polluted. Polluting and damaging the rivers, forests, lakes, water bodies, air and glaciers will be legally equivalent to harming, hurting and causing injury to person. Rivers, Forests, Lakes, Water Bodies, Air, Glaciers and Springs have a right to exist, persist, maintain, sustain and regenerate their own vital ecology system. The rivers are not just water bodies. These are scientifically and biologically living…Rivers are grasping for breath. We must recognize and bestow the Constitutional legal rights to the Mother Earth".

Further, because of the ecological interdependence of different natural entities, the court deemed it necessary to grant personhood to a wide variety of entities, declaring "Glaciers including Gangotri & Yamunotri, rivers, streams, rivulets, lakes, air, meadows, dales, jungles, forests wetlands, grasslands, springs and waterfalls, legal entity/legal person/juristic person/juridicial person/moral person/artificial person having the status of a legal person, with all corresponding rights, duties and liabilities of a living person, in order to preserve and conserve them. They are also accorded the rights akin to fundamental rights/legal rights." The Indian Supreme Court later stayed the Ganges and Yamuna decision, not because it disagreed with the ability of the court to create these legal persons, but because of concerns over the potential liability of the state if these entities should be sued—that is, how nature's duties would be interpreted (*40*). Although this decision was stayed, the reasoning has not been overturned and has been influential in some other cases.

Judicial refusal to engage with science in the USA

Drewes Farms Partnership v. *Toledo*

The rights of Lake Erie and the Lake Erie Watershed to "exist, flourish, and naturally evolve" were enacted through a ballot initiative in Toledo, Ohio. This law was challenged in court by a farm company and the state of Ohio. In Drewes Farms Partnership versus Toledo, the court refused to contemplate what the new rights meant at all in its decision to invalidate the Lake Erie Bill of Rights.

This law was one of many local laws in the USA that purported to invalidate permits granted by state or other authorities that

would violate nature's rights and to invalidate the legal personhood of corporations that would violate nature's rights. These provisions conflict with US jurisprudence on federalism and corporate rights, and the court unsurprisingly struck down the law. However, the court decided the case on other grounds, ruling that the rights to "exist, flourish, and naturally evolve" were unconstitutionally vague. In its order invalidating the law, the court asked "[w]hat conduct infringes the right of Lake Erie and its watershed to 'exist, flourish, and naturally evolve'? How would a prosecutor, judge, or jury decide?" This was surprising in that courts commonly do adjudicate matters with scientific aspects. The court did not discuss any scientific or other sources available that could have been used to help interpret these terms. This decision potentially calls into question the legal system's ability to adjudicate scientific matters, or at least courts' willingness to do so in novel situations. On the other hand, this decision points to a way forward for rights of nature advocates: writing clearer or more scientifically descriptive laws may make it more likely that the laws will be upheld and given their intended effect by courts.

The right to evolve

Rights-of-nature laws and court decisions necessarily use many scientific concepts. One that recurs in many rights-of-nature laws is evolution and the "right to evolve" (Table 1). Few courts have yet interpreted this right, which makes scientific input as to how it might be interpreted even more urgent. Further, interdisciplinary examination of how evolution might be interpreted in different contexts may help drafters of proposed rights-of-nature laws in formulating this right. In this section we highlight some of the difficulties in understanding how this right applies to different types of natural entities and how it can be given effect. We then suggest how evolutionary biology can contribute to more clearly defining this new legal right.

Although scientific concepts are used widely in law, their scientific meaning, legal usage, and popular understanding rarely coincide precisely. Further, as many scholars have noted, scientific concepts and determinations are often contingent on social values to varying extents and contain varying levels of uncertainty; therefore recourse to science does not necessarily enable judges to apply objectively correct answers to legal questions (*41*, *42*, *43*, *44*, *45*). Legal decision-makers frequently must choose among differing scientific interpretations or interpret terms that have different meanings in different contexts. For instance, the US Supreme Court famously ruled in 1893 that tomatoes are legally considered vegetables, at least for the purpose of an import tax (*46*). The court acknowledged that botanically tomatoes are fruit, but in the common language of the people they are vegetables and the common usage of tomatoes in salads rather than desserts was of decisive importance. In a contrasting legal context, when tomatoes are used in jams the EU Directive on fruit jams, jellies and marmalades, and sweetened chestnut purée classifies them as fruit (*47*). Although this classification agrees with the scientific definition of a fruit, it also reflects the popular idea of jams being made from fruit.

Evolution, like tomatoes, has many possible definitions. Although a broadly understood definition is something like "slow change over time," this is only the 7th (of 11) definitions of evolution in the Oxford English Dictionary (OED) (*48*). Number eight in the OED is the Darwin-inspired definition: "the transformation of animals, plants, and other living organisms into different forms by the accumulation of changes over successive generations; the transmutation of species." The other definitions include usage in dance, gymnastics, mathematics, chemistry, the military, and development (of societies, institutions, systems, products, arguments, organisms, and organs). There is a real possibility that different definitions of "evolve" in the "right to evolve" or "right to evolutionary processes" will be used in different legal jurisdictions. Further, many rights-of-nature laws specify the rights of lakes, ecosystems, and other entities to evolve. Group selection and evolution of nonhuman communities or ecosystems, though not uncontroversial, are recognized by some ecologists and evolutionary biologists (*49*), but this kind of evolution is not even included within the OED's definitions of the term.

Any or all these definitions can contribute to understanding what a "right to evolve" may entail, whether for a troop of monkeys, a population of wild rice, Lake Erie, or Pachamama. For example, a biological understanding of evolution may be applied to species or subspecies, whereas a more general "slow change over time" may pertain to rivers, lakes, and watersheds. The simultaneous application of multiple definitions of evolution is illustrated by the September 2021 decision of Ecuador's Constitutional Court regarding the rights of mangrove ecosystems (*50*).

In that case, the Court confirmed "that mangrove ecosystems are holders of the recognized rights of nature" and are entitled to "full respect for their existence and the maintenance and regeneration of their life cycles, structure, functions and evolutionary processes." As the Court noted, mangrove ecosystems are marine-coastal wetlands that provide habitat to many animal species and contribute to mitigating climate change. Explaining what it might mean to disrupt the evolutionary process of mangroves, the Court stated that "natural beings respond to long processes of permanent changes that allow adaptation to the environment. The rupture of the elements that allow an evolutionary process would constitute a violation of the rights of nature." The court ruling implied that mangroves not only had the right to their evolutionary processes as species, which would require sufficient habitat and genetic diversity, but also had rights as ecosystems, implying that other species that relied on the mangroves as habitat should also have sufficient habitat and genetic diversity to evolve. This is consistent with current evolutionary thinking, which asserts that evolution is a property of populations, not individuals, and therefore requires a sustainable population size and sufficient within-population genetic diversity to enable adaptation to a changing environment (*51*, *52*).

We note that the right to evolve may conflict with the right of a natural entity to exist or persist; over the history of Earth, change and extinction is the rule, not the exception: >99% of all species that have evolved in the last billion years have gone extinct (*53*). We are unaware of a rights-of-nature law that deals effectively with this contradiction or a challenge to a rights-of-nature law based on this conflict, but several cases have placed more emphasis on the right to exist rather than the right to evolve. Regardless, the timescales of existence, evolution, and natural extinction are potentially much longer than the human timescales that courts are used to dealing with and present an additional challenge for courts in making decisions that protect the rights of nature. This novelty makes scientific input even more important in enabling the legal system to meaningfully adjudicate rights of nature.

Although science alone cannot define what nature's rights mean in the legal systems that have incorporated them, science is a necessary component for understanding, for instance, evolutionary processes, ecological flows, speciation, extinction, and natural cycles, if and when nature or other entities are given rights to these processes.

Conclusions

The recent legal developments discussed in this Review demonstrate the importance of scientists to the implementation of rights-of-nature laws. The Constitutional Court of Ecuador has held that various entities, including the Los Cedros Forest and mangrove ecosystems, were rights holders, and that mining and fishing violated their rights. The testimony of ecologists, biologists and other scientists was essential to the Court's understanding of who these rights holders were and what industrial activities had violated their rights to their ecological functions. By contrast, other rights-of-nature laws have been held to be unconstitutional by courts, at least in one case because of the challenge of using scientific knowledge to protect rights. The 2019 Lake Erie Bill of

Rights, which overwhelmingly passed by a popular vote in Toledo, Ohio, was later held to be unconstitutionally vague on the grounds that it would be impossible to determine when Lake Erie's rights to "exist, flourish, and naturally evolve" were violated (*54*). Spain's Mar Menor law, though more detailed as to the rightsholder and rights than the Lake Erie law, may be headed for the same fate. The Spanish right-wing party Vox has challenged this law in Spain's constitutional court, alleging that it is unconstitutionally vague among other claimed shortcomings (*55*).

Because legal decision-makers often do not have the expertise to understand the rights and duties of nature, interdisciplinary research is urgently needed if rights-of-nature laws are to have legal meaning. By contributing to interdisciplinary analyses of different aspects of rights-of-nature laws before disputes arise, scientists can help contribute to the understanding and effectiveness of these laws. Although scientific uncertainty often cannot be eliminated (*4*), it's reduction in turn reduces legal uncertainty and thus helps meet the objection by some judges and others that rights-of-nature laws are too vague to be applied.

There are many ways that scientists and scientific knowledge can help different legal systems understand nature's rights. One important way scientists can contribute is by being involved in litigation. Most rights-of-nature laws contain provisions allowing scientists or other members of the public to bring lawsuits. A successful lawsuit necessitates demonstration that the entity in question is protected and that its rights were violated—e.g., how its rights to ecological functions or evolutionary processes were impeded. These are clearly questions requiring scientific input and applications of ecological and evolutionary concepts. Scientists also can provide evidence in lawsuits to which they are not a party in many legal systems. Amicus curiae briefs and testimony from scientists were critically important in the Ecuadorian cases.

Many countries also allow the public, including scientists, to participate in environmental decision-making. Participation of scientists additionally may be built into the legal recognition of rights for nature: Grants of rights to natural entities often specify a process for ensuring that their rights are achieved, and many of these processes explicitly include a role for scientists. For example, the Mar Menor law established a scientific committee made up of "scientists and independent experts specialized in the study of the Mar Menor" (*30*). The scientists included in this committee are appointed, but these sorts of committees are another opportunity for scientists with relevant knowledge about particular rights-holding entities to be involved in understanding and upholding their rights.

Critically, even before a natural entity's rights are violated, scientists and legal scholars can help lay the scientific groundwork for these legal protections to function as intended. As rights-of-nature laws are new and many are untested, interdisciplinary scholarship applying both legal and scientific perspectives to explain concepts used in the legal protection of nature's rights is needed for the legal system to correctly adjudicate these rights. The availability of credible scholarly analysis of legal-scientific terms used in law would make these rights more tangible and accessible to the judges whose role it is to apply them.

Such scholarship can clarify what aspects of the laws require scientific input, and what aspects require other types of expertise. For example, the Mar Menor lagoon has the rights to be restored and to exist as a lagoon ecosystem. A legal scholar may use legal sources to explain to what level of restoration Mar Menor is legally entitled. However, it may not be clear from legal sources who decides the restoration baseline, and scientific input may also be necessary. A restoration ecologist could explain what may be required to achieve the required level of restoration but what measures are legally appropriate may also require input from social scientists and others, as humans are also considered to be part of the protected ecosystem.

Interdisciplinary scholarship is also of benefit to the many jurisdictions that are currently considering enacting rights-of-nature laws (*56*). As we have illustrated, the content of many of the terms used in rights-of-nature laws is unclear. Even though the legal definition of a term need not be the same as a scientific definition, scholarship exploring these concepts may help drafters of rights-of-nature laws choose terms that minimize conflicts between legal and scientific contexts, enabling new laws to be easier to apply and follow.

Another type of interdisciplinary scholarship that would assist the functioning of rights-of-nature laws would be the examination of the duties of nature. Although some of these laws, such as those in the Ecuador constitution, only grant rights for nature without corresponding duties, others, such as the New Zealand laws, equate nature to a legal person with both rights and duties. Indeed, uncertainty over liabilities and duties of nature was an impediment to implementing the Indian court decision assigning certain rivers legal personhood (*2*). Input from scientists is needed to understand what duties nature has (e.g., "ecosystem services") (*57*) and when nature may not be fulfilling these duties. By delineating, for example, the volume and quality of water provided by a river, scientists can help legal systems comprehend nature's potential legal obligations and understand what environmental protection measures may also be legally required to ensure natural entities can continue to fulfil these obligations.

REFERENCES AND NOTES

1. IPBES, "Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services," S. Díaz *et al*., Eds. (IPBES secretariat, 2019); https://doi.org/10.5281/zenodo.3553579.doi: 10.5281/zenodo.3553579
2. C. M. Kauffman, P. L. Martin, *The Politics of Rights of Nature* (MIT Press, 2021).
3. E. Fisher, in *The Oxford Handbook of Comparative Environmental Law* (Oxford Univ. Press, 2019) pp. 749–768.
4. E. Lees, T. Paloniitty, in *EU Environmental Principles and Scientific Uncertainty Before National Courts: The Case of the Habitats Directive.* (Hart Publishing, 2023) pp. 9–30.
5. O. Houck, Tales from a Troubled Marriage: Science and Law in Environmental Policy. *Science* **302**, 1926–1929 (2003). doi: 10.1126/science.1093758
6. E. O'Donnell, R. Nelson, Shield Science for Robust Decisions. *Nat. Sustain* **3**, 675–676 (2020). doi: 10.1038/s41893-020-0541-6
7. L. Wahlberg, Legal Ontology, Scientific Expertise and The Factual World. *Journal of Social Ontology* **3**, 49–65 (2017). doi: 10.1515/jso-2015-0022
8. C. Case, 528/16. ECLI:EU C, 583 (2018).
9. *County of Maui* v. *Hawaii Wildlife Fund*, 140 S. Ct. 1462 (2020).
10. B. Mayer, The Contribution of Urgenda to the Mitigation of Climate Change.*Journal of Environmental Law*eqac016 (2022). doi: 10.1093/jel/eqac016
11. M. J. Currell, D. J. Irvine, A. D. Werner, C. McGrath, Science sidelined in approval of Australia's largest coal mine. *Nat. Sustain.* **3**, 644–649 (2020). doi: 10.1038/s41893-020-0527-4
12. M. V. Berros, Challenges for the Implementation of the Rights of Nature: Ecuador and Bolivia as the First Instances of an Expanding Movement.*Latin American Perspectives* 192–205 (2021). doi: 10.1177/0094582X211004898
13. A. F. Chalmers, *What Is This Thing Called Science?* (Open Univ. Press, 1982) (Second edition).
14. J. Needham, *Science in Traditional China: A Comparative Perspective.* (Harvard Univ. Press, 1981).
15. J. Clapcott *et al*., Mātauranga Māori: Shaping Marine and Freshwater Futures. *N. Z. J. Mar. Freshwater Res.* **52**, 457–466 (2018). doi: 10.1080/00288330.2018.1539404
16. V. Watts, Indigenous place-thought & agency amongst humans and non-humans (First Woman and Sky Woman go on a European world tour!). *Decolonization: Indigeneity, Education & Society.* **2**, 20–34 (2013).
17. M. Muru-Lanning, Mātauranga Māori Science and the Appropriation of Water in New Zealand. *Anthropol. Forum.* **22**, 151–164 (2012). doi: 10.1080/00664677.2012.694171
18. E. O'Donnell, A. Poelina, A. Pelizzon, C. Clark, Stop Burying the Lede: The Essential Role of Indigenous Law(s) in Creating Rights of Nature. *Transnational Environmental Law* **9**, 403–427 (2020). doi: 10.1017/S2047102520000242
19. M. Tănăsescu, *Understanding the Rights of Nature: A Critical Introduction*. (Transcript verlag, 2022).
20. J. R. Leaf Sr., What is Traditional Ecological Knowledge and why does it matter? *Frontiers in Ecology and the Environment* **20**, 3 (2022). doi: 10.1002/fee.2465
21. G. Chapron, Y. Epstein, J. López-Bao. A rights revolution for nature. *Science* **363**, 1392–1393 (2019). doi: 10.1126/science.aav5601
22. C. M. Kauffman, P. L. Martin, Constructing Rights of Nature Norms in the US, Ecuador, and New Zealand. *Global Environmental Politics* **18**, 43–62 (2018). doi: 10.1162/glep_a_00481
23. Constitucion de la Republica del Ecuador; 2008.
24. E. O'Donnell, *Legal Rights for Rivers: Competition, Collaboration and Water Governance.* (Routledge, 2018).
25. Te Awa Tupua (Whanganui River Claims Settlement) Act 2017, No. 17. (2017).
26. J. Talbot-Jones, J. Bennett, Implementing bottom-up governance through granting legal rights to rivers: a case study of the Whanganui River, Aotearoa New Zealand. *Australas. J. Environ. Manag* **29**, 64–80 (2022). doi: 10.1080/14486563.2022.2029775
27. E. O'Donnell, Rivers as living beings: rights in law, but no rights to water? *Griffith Law Review* **29**, 643–668 (2021). doi: 10.1080/10383441.2020.1881304
28. Panama, Ley No. 287 que reconoce los derechos de la naturaleza y las obligatciones del estado relacionadas con estos derechos (Law No. 287, 2022).

29. E. Stokstad, This lagoon is effectively a person, says Spanish law that's attempting to save it. *Science* **378**, 15–16 (2022). doi: 10.1126/science.adf1102
30. Spain, Ley 19/2022, de 30 de septiembre, para el reconocimiento de personalidad jurídica a la laguna del Mar Menor y su cuenca (Law 19/2022, 30 September on the recognition of the legal personality of the Mar Menor lagoon and its basin). 2022.
31. An Ordinance of the City Council of the City of Santa Monica Establishing Sustainability Rights (2013); https://library.qcode.us/lib/santa_monica_ca/pub/municipal_code/item/article_12-chapter_12_02-12_02_030.
32. White Earth Reservation Business Committee Resolution Nos. 001-19-009 and 001-19-010; The 1855 Treaty Authority's Resolution No. 2018-05.
33. E. Fitz-Henry, Challenging Corporate "Personhood": Energy Companies and "Rights" of Non-Humans. *Political and Legal Anthropology Review* **41**, 85–102 (2018). doi: 10.1111/plar.12255
34. Santa Monica Sustainable Groundwater Management Ordinance, Ord. No. 2584CCS § 1. (2018).
35. M. Immovilli, S. Reitsma, R. Roncucci, E. D. Rasch, Exploring Contestation in Rights of River Approaches: Comparing Colombia, India and New Zealand. *Water Alternatives* **15**, 547–591 (2022).
36. *Corte Constitucional del Ecuador* (Constitutional Court of Ecuador), case 2167-21-EP (2022).
37. *Corte Constitucional del Ecuador* (Constitutional Court of Ecuador), case 32-17-IN (2021).
38. *Corte Constitucional del Ecuador* (Constitutional Court of Ecuador), case 1149-19-JP/20 (2021).
39. E. Mayr, *Systematics and the Origin of Species.* (Columbia Univ. Press, 1942).
40. E. O'Donnell, At the Intersection of the Sacred and the Legal: Rights for Nature in Uttarakhand, India. *J. Environ. Law* **30**, 135–144 (2018). doi: 10.1093/jel/eqx026
41. B. Latour, *Pandora's Hope: Essays on the Reality of Science Studies.* (Harvard Univ. Press, 1999).
42. H. Putnam, *The Collapse of the Fact/Value Dichotomy.* (Harvard Univ. Press, 2002).
43. S. Jasanoff, *Science and Public Reason.* (Routledge, 2012).
44. J. Preston, Biodiversity in the Court: The Certainty of Contests about Uncertainty. M. Eliantonio, E. Lees, T. Paloniitty, Eds., *EU Environmental Principles and Scientific Uncertainty before National Courts* (2023).
45. Y. Epstein, Favourable Conservation Status for Species: Examining the Habitats Directive's Key Concept through a Case Study of the Swedish Wolf. *J. Environ. Law* **28**, 221–244 (2016). doi: 10.1093/jel/eqw026
46. *Nix* v. *Hedden.* 149 U.S. 304, 1893.
47. Council Directive 2001/113/EC of 20 December 2001 relating to fruit jams, jellies and marmalades and sweetened chestnut purée intended for human consumption.
48. *Oxford English Dictionary Online*, (Oxford Univ. Press, 2022).
49. E. G. Leigh Jr., The Group Selection Controversy. *J. Evol. Biol.* **23**, 6–19 (2010). doi: 10.1111/j.1420-9101.2009.01876.x
50. *Corte Constitucional del Ecuador* (Constitutional Court of Ecuador), case 22-18-IN (2021).
51. T. J. Kawecki *et al*., Experimental Evolution. **27**, P547–560 (2012). doi: 10.1016/j.tree.2012.06.001
52. B. R. Forester, E. A. Beever, C. Darst, J. Szymanski, W. C. Funk, Linking Evolutionary Potential to Extinction Risk: Applications and Future Directions. *Front Ecol Environ* **20**, 493–544 (2022). doi: 10.1002/fee.2552
53. C. R. Marshall, Five Palaeobiological Laws Needed to Understand the Evolution of the Living Biota. *Nat. Ecol. Evol* **1**, 0165 (2017). doi: 10.1038/s41559-017-0165
54. *Drewes Farms P'ship* v. *City of Toledo*, 441 F.Supp.3d 551, 2020.
55. D. Gomez, Vox lleva al Tribunal Constitucional la ley de personalidad jurídica del Mar Menor. (La Verdad, 2023); https://www.laverdad.es/murcia/lleva-tribunal-constitucional-20230110124310-nt.html.
56. A. Putzera, T. Lambooyb, R. Jeurissenband, E. Kim, Putting the rights of nature on the map. A quantitative analysis of rights of nature initiatives across the world. *Journal of Maps* **18**, 89–96 (2022). doi: 10.1080/17445647.2022.2079432
57. Millennium Ecosystem Assessment, Ecosystems and Human Well-being: Synthesis (Island Press, 2005).

ACKNOWLEDGMENTS

Two anonymous reviewers provided helpful comments and critiques of an early version of the manuscript. We thank the Swedish Collegium for Advanced Study and Riksbankens Jubileumsfond for providing funding and infrastructure that facilitated this research. We also thank Formas for funding project 019-01450. We note that when non-English texts have been quoted, we have used official translations where available, and otherwise used non-official or our own translations. **Competing Interests:** H.E. works as external attorney in Ecuador for the American nonprofit organization Center for Democratic and Environmental Rights (CDER), which advocates for rights of nature. He contributed to this article in his role as faculty member at Universidad Hemisferios of Quito, Ecuador. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

Submitted 4 November 2022; accepted 6 April 2023
10.1126/science.adf4155

RESEARCH ARTICLE SUMMARY

HUMAN GENOMICS

Discovery of target genes and pathways at GWAS loci by pooled single-cell CRISPR screens

John A. Morris, Christina Caragine, Zharko Daniloski, Júlia Domingo, Timothy Barry, Lu Lu, Kyrie Davis, Marcello Ziosi, Dafni A. Glinos, Stephanie Hao, Eleni P. Mimitou, Peter Smibert, Kathryn Roeder, Eugene Katsevich, Tuuli Lappalainen*, Neville E. Sanjana*

INTRODUCTION: Genome-wide association studies (GWASs) have identified thousands of human genetic variants associated with diverse diseases and traits, and most of these variants map to noncoding loci with unknown target genes and function. Current approaches to understand which GWAS loci harbor causal variants and to map these noncoding regulators to target genes suffer from low throughput. With newer multi-ancestry GWASs from individuals of diverse ancestries, there is a pressing and growing need to scale experimental assays to connect GWAS variants with molecular mechanisms.

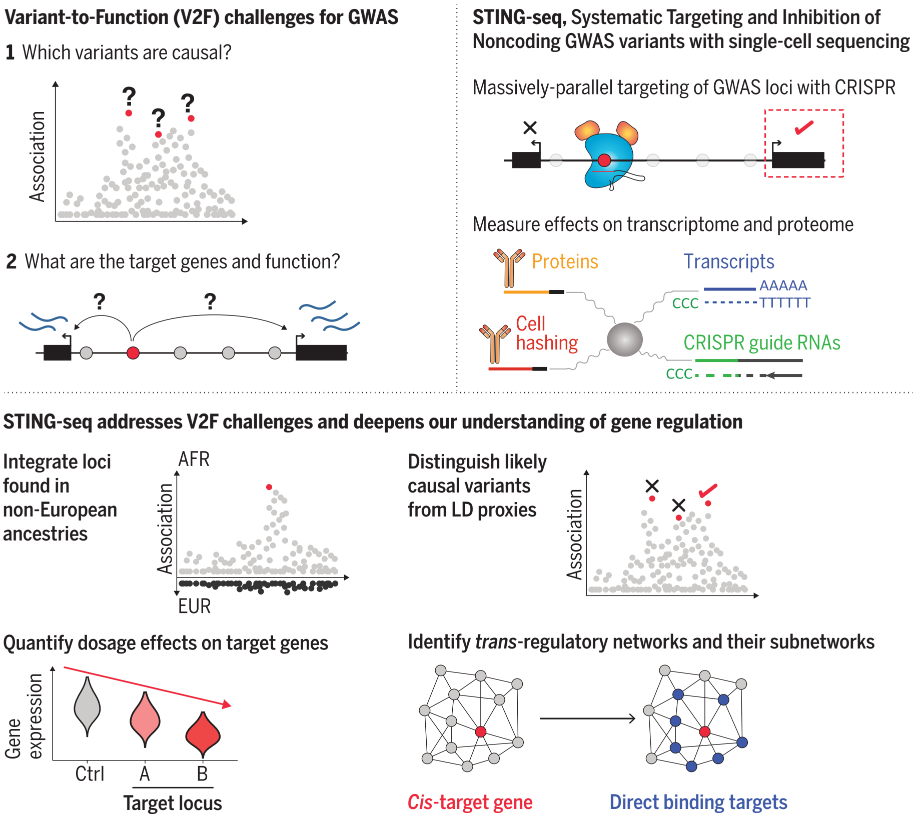
Here, we combined biobank-scale GWASs, massively parallel CRISPR screens, and single-cell sequencing to discover target genes of noncoding variants for blood trait loci with systematic targeting and inhibition of noncoding GWAS loci with single-cell sequencing (STING-seq).

RATIONALE: Blood traits are highly polygenic, and GWASs have identified thousands of noncoding loci that map to candidate *cis*-regulatory elements (CREs). By combining CRE-silencing CRISPR perturbations and single-cell readouts, we targeted hundreds of GWAS loci in a si assay, revealing target genes in *cis* and in *tr*. For select CREs that regulate target genes, we performed direct variant insertion. Although silencing the CRE can identify the target gene, direct variant insertion can identify magnitude and direction of effect on gene expression for the GWAS variant. In select cases in which the target gene was a transcription factor or microRNA, we also investigated the gene-regulatory networks altered upon CRE perturbation and how these networks differ across blood cell types.

RESULTS: We inhibited candidate CREs from fine-mapped blood trait GWAS variants (from ~750,000 individual of diverse ancestries) in human erythroid progenitors. In total, we targeted 543 variants (254 loci) mapping to candidate CREs, generating multimodal single-cell data including transcriptome, direct CRISPR gRNA capture, and cell surface proteins.

We identified target genes in *cis* (within 500 kb) for 134 CREs. In most cases, we found that the target gene was the closest gene and that specific enhancer-associated biochemical hallmarks (H3K27ac and accessible chromatin) are essential for CRE function. Using multiple perturbations at the same locus, we were able to distinguish between causal variants from noncausal variants in linkage disequilibrium. For a subset of validated CREs, we also inserted specific GWAS variants using base-editing STING-seq (beeSTING-seq) and quantified the effect size and direction of GWAS variants on gene expression. Given our transcriptome-wide data, we examined dosage effects in *cis* and *trans* in cases in which the *cis* target is a transcription factor or microRNA. We found that *trans* target genes are also enriched for GWAS loci, and identified gene clusters within *trans* gene networks with distinct biological functions and expression patterns in primary human blood cells.

CONCLUSION: In this work, we investigated noncoding GWAS variants at scale, identifying target genes in single cells. These methods can help to address the variant-to-function challenges that are a barrier for translation of GWAS findings (e.g., drug targets for diseases with a genetic basis) and greatly expand our ability to understand mechanisms underlying GWAS loci. ■



Identifying causal variants and their target genes with STING-seq. Uncovering causal variants and their target genes or function are a major challenge for GWASs. STING-seq combines perturbation of noncoding loci with multimodal single-cell sequencing to profile hundreds of GWAS loci in parallel. This approach can identify target genes in *cis* and *trans*, measure dosage effects, and decipher gene-regulatory networks.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: tlappalainen@nygenome.org (T.L.); neville@sanjanalab.org (N.E.S.)
Cite this article as J. A. Morris *et al.*, *Science* **380**, eadh7699 (2023). DOI: 10.1126/science.adh7699

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.adh7699

RESEARCH ARTICLE

HUMAN GENOMICS

Discovery of target genes and pathways at GWAS loci by pooled single-cell CRISPR screens

John A. Morris[1,2], Christina Caragine[1,2], Zharko Daniloski[1,2], Júlia Domingo[1], Timothy Barry[3], Lu Lu[1,2], Kyrie Davis[1,2], Marcello Ziosi[1], Dafni A. Glinos[1], Stephanie Hao[4], Eleni P. Mimitou[4], Peter Smibert[4], Kathryn Roeder[3,5], Eugene Katsevich[6], Tuuli Lappalainen[1,7]*, Neville E. Sanjana[1,2]*

Most variants associated with complex traits and diseases identified by genome-wide association studies (GWAS) map to noncoding regions of the genome with unknown effects. Using ancestrally diverse, biobank-scale GWAS data, massively parallel CRISPR screens, and single-cell transcriptomic and proteomic sequencing, we discovered 124 *cis*-target genes of 91 noncoding blood trait GWAS loci. Using precise variant insertion through base editing, we connected specific variants with gene expression changes. We also identified *trans*-effect networks of noncoding loci when *cis* target genes encoded transcription factors or microRNAs. Networks were themselves enriched for GWAS variants and demonstrated polygenic contributions to complex traits. This platform enables massively parallel characterization of the target genes and mechanisms of human noncoding variants in both *cis* and *trans*.

A major goal for the study of common diseases is to identify causal genes, which can clarify biological mechanisms and inform drug targets for these diseases. To this end, genome-wide association studies (GWASs) have identified thousands of genetic variants associated with disease outcomes and disease-relevant phenotypes. However, because these associations are nearly always found in noncoding regions, their target genes and functions often remain elusive. This is commonly referred to as the variant-to-function (V2F) problem (*1*, *2*).

Recent studies have used statistical fine-mapping to identify plausibly causal GWAS variants and functional genomics to find candidate *cis*-regulatory elements (cCREs) and their putative target genes (*3*–*6*). Other studies have performed CRISPR-based silencing or mutagenesis screens of noncoding regulatory elements to identify target genes (*7*–*9*). Here, we combined these approaches in a modular workflow, systematic targeted inhibition of noncoding GWAS loci coupled with single-cell sequencing (STING-seq), to identify target genes at noncoding GWAS loci using single-cell pooled CRISPR screens. We first prioritized cCREs by functional annotation and overlap with fine-mapped GWAS variants. We then tested for gene-regulatory function using pooled CRISPR inhibition (CRISPRi) and single-cell RNA-sequencing and cell surface protein measurements (Fig. 1A). For a subset of validated CREs, we also inserted specific GWAS variants using base editing STING-seq (beeSTING-seq), which couples base editing with single-cell multiomics. We demonstrate the utility of these approaches in blood cell traits by targeted perturbation of ~500 cCREs at noncoding GWAS loci, identifying target genes in *cis* and *trans* for 134 of these CREs, and further explore the effects of 46 fine-mapped noncoding C-to-T variants using precise variant insertion.

[1]New York Genome Center, New York, NY 10013, USA. [2]Department of Biology, New York University, New York, NY 10003, USA. [3]Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA. [4]Technology Innovation Lab, New York Genome Center, New York, NY 10013, USA. [5]Computational Biology Department, Carnegie Mellon University, Pittsburgh, PA 15213, USA. [6]Department of Statistics and Data Science, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104, USA. [7]Science for Life Laboratory, Department of Gene Technology, KTH Royal Institute of Technology, 171 65 Solna, Stockholm, Sweden.
*Corresponding author. Email: tlappalainen@nygenome.org (T.L.); neville@sanjanalab.org (N.E.S.)

Results

Fine-mapping multiancestry blood trait GWAS to identify candidate CREs

We elected to study blood cell traits because of their high polygenicity, links to multiple common diseases, and the large number of genotyped individuals available in ancestrally diverse biobank-scale data repositories with measured blood traits (*10*–*12*). We examined 29 blood trait GWASs in the UK Biobank (UKBB) and 15 traits from the Blood Cell Consortium (BCX) (*11*), including traits from platelets, red blood cells (RBCs), and white blood cells (WBCs) (table S1A). The UKBB GWASs include 361,194 participants with European ancestries. The BCX multiancestry GWASs include 746,667 participants (76% European, 20% Asian, 2% African, 1% Hispanic/Latino, and 1% South Asian ancestries) with both multiancestry and individual population analyses. We performed statistical fine-mapping for the 29 UKBB blood trait GWASs, identifying a median of 469 conditionally independent signals and 3328 fine-mapped variants per trait (*13*, *14*). Multiancestry BCX meta-analyses identified a median of 384 conditionally independent signals and 3586 fine-mapped variants per trait. Across all BCX population-specific GWASs, excluding European ancestries, there were 42 conditionally independent signals and 418 fine-mapped variants per trait (table S1, A and B). In all cases, we found that >90% of fine-mapped variants were in noncoding regions of the genome.

For our study, we targeted cCREs from different GWASs (543 variants in 254 loci) by intersecting fine-mapped noncoding variants with biochemical hallmarks of enhancer activity, such as chromatin accessibility [assay for transposase-accessible chromatin sequencing (ATAC-seq) and DNase I hypersensitivity] and canonical histone modifications [H3K27ac chromatin immunoprecipitation sequencing (ChIP-seq)] from the human erythroid progenitor cell line K562. K562 cells are an established and well-characterized model for blood traits. In these cells, reporter assays have identified genetic variants with erythroid-specific effects (*15*), transcription factor (TF) occupancy is strongly conserved with human proerythroblasts (*16*), gene expression and open chromatin profiles are similar to human erythrocyte progenitors (*17*), and promoter-interacting regions defined from Hi-C data are enriched for blood trait GWAS variants (*18*). The integration of functional genomic data yielded a large set of targetable variants from UKBB and BCX GWASs (table S1, C and D). The variants that we selected were often the highest-probability variant in a fine-mapped GWAS locus (294 variants) or among the 10 most probable variants (249 variants). We also prioritized variants from non-European ancestries. In total, we selected variants from BCX multiancestry analyses (339 variants), BCX non-European ancestries (118 variants), and UKBB European ancestries (86 variants) (Fig. 1B and table S1, C to E).

Optimized dual-repressor CRISPRi system

To perturb the selected cCREs, we designed (table S1F) a dual-repressor KRAB-dCas9-MeCP2 system (*19*) that yielded 50 to 60% greater gene repression when targeting transcription start sites (TSSs) or previously described enhancer loci (*7*) than a single-repressor (KRAB-dCas9) system (Fig. 1, C and D; fig. S1; and table S2). We further characterized the dual-repressor CRISPRi using a pooled library of ~2000 CRISPR guide RNAs (gRNAs) that target sites at different distances from the TSSs of ~250 essential genes. We found that dual-repressor CRISPRi had a focused activity window with minimal repression beyond 1 kb, and that most of the active gRNAs were located between –400 and +850 nucleotides (nt) from the TSS (fig. S2) (*20*).

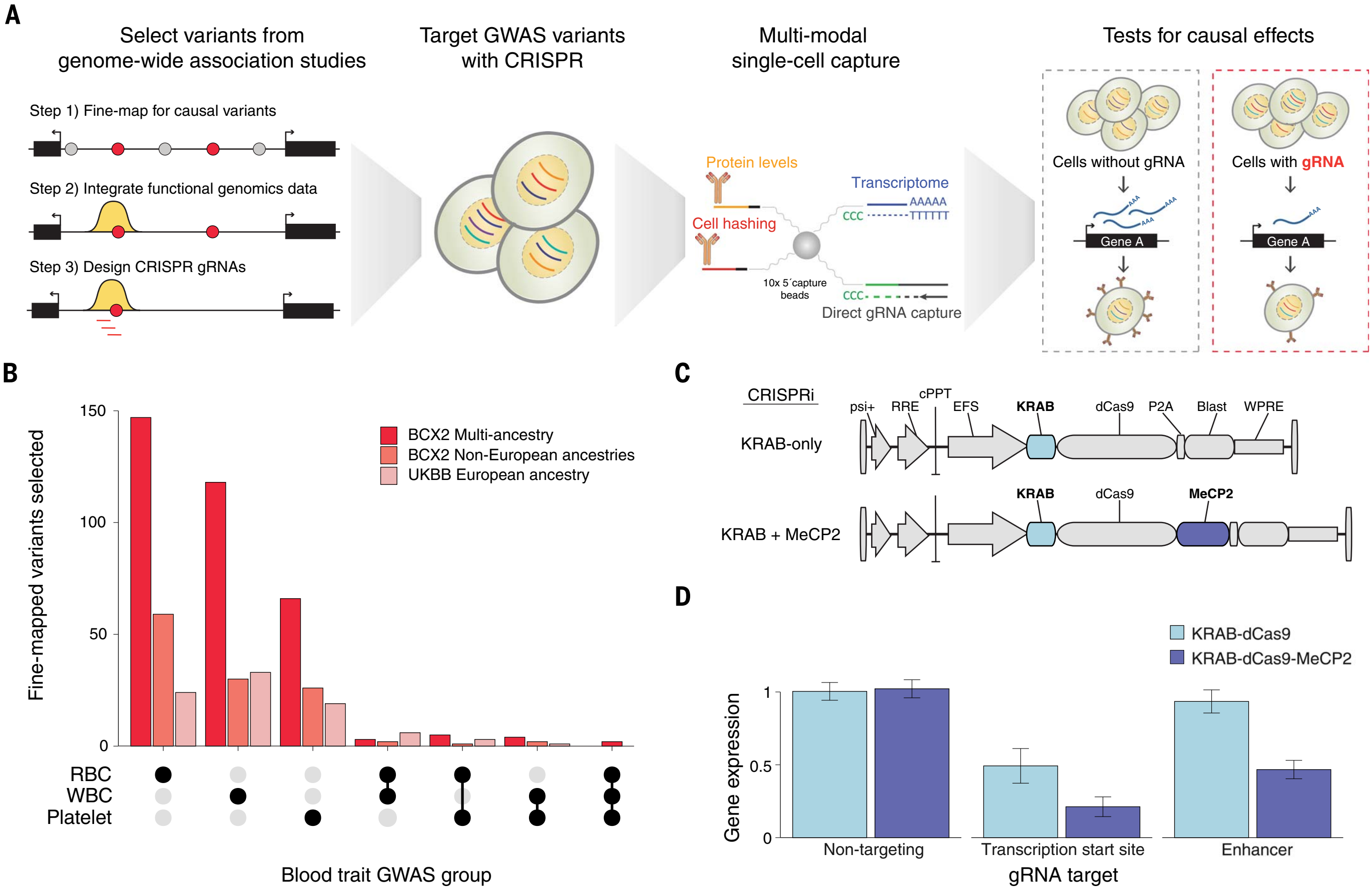


Fig. 1. Overview of STING-seq. (**A**) STING-seq pipeline for perturbation and single-cell analysis of human genetic variants from GWASs. First, plausibly causal variants are identified using statistical fine-mapping of GWAS. After further refinement of candidate cCREs using key molecular hallmarks of regulatory elements, CRISPR gRNAs are designed to target cCREs and lentivirally transduced at a high MOI into human cells. Using multimodal single-cell sequencing, target genes for GWAS variants are identified using differential transcript or protein expression. (**B**) The number of targeted GWAS variants mapping to cCREs across 29 blood traits in UKBB (n = 361,194 participants) and 15 blood traits in the BCX (n = 746,667 participants). (**C**) Lentiviral CRISPRi vector with a single-effector domain (KRAB-dCas9) or dual-effector domains (KRAB-dCas9-MeCP2). (**D**) Mean digital PCR gene expression in human erythrocyte cells (K562) by targeting the TSSs and known enhancers of three genes (*MRPS23*, *SLC25A27*, and *FSCN1*) with either single-effector KRAB-dCas9 or dual-effector KRAB-dCas9-MeCP2 CRISPRi. Error bars indicate SEM.

A massively parallel assay to perturb CREs and find their target genes

We designed STING-seq gRNA libraries to target each blood trait cCRE with up to three gRNAs using the dual-repressor CRISPRi (KRAB-dCas9-MeCP2). These gRNAs were optimized for minimal off-target activity (*21*, *22*). We also embedded into the STING-seq library several control gRNAs: negative (nontargeting) controls (*23*), positive controls (targeting highly-expressed genes at TSSs), and, to estimate the average number of perturbations per cell through flow cytometry, multiple gRNAs targeting a gene encoding a ubiquitously expressed cell surface protein (*CD55*) (table S3A).

We transduced K562 cells with pooled library virus at a high multiplicity of infection (MOI), which we verified by flow cytometry for CD55 (fig. S3). We then simultaneously captured four different modalities from single cells: CRISPR gRNAs, transcriptomes, cell surface proteomes through oligo-tagged antibodies, and cell hashing (table S3B) (*24*, *25*). We recovered 46,583 single cells with a median of 13 gRNAs per cell and with each cCRE targeted in a median of 978 cells (fig. S4, A and B, and table S3C). To perform differential expression testing, we recently developed a conditional resampling approach (SCEPTRE) that yields state-of-the-art calibration on CRISPR single-cell datasets to connect perturbations with changes in gene and protein expression (*26*). Using SCEPTRE, we grouped together gRNAs targeting each cCRE, performing 4627 pairwise tests with a median of seven genes tested per cCRE within 500 kb for *cis* effects (*27*). We observed good calibration for positive and negative controls: Nontargeting gRNAs had no effect, and control genes had decreased expression or protein levels at a 5% false discovery rate (FDR) (Fig. 2A; fig. S5; and table S3, C to E). In most cases, target genes in *cis* for GWAS variants were more likely to be identified when both H3K27ac and open chromatin peaks were present (Fig. 2B).

Of 539 targeted cCREs (from 254 loci), 134 (from 91 loci) had a target gene within 500 kb (Fig. 2C and table S3F). When examining gRNAs that target the same CRE, the number of cells was most directly responsible for statistical power, and not distance between gRNAs or predicted off-target effects (fig. S6). We found minimal differences in target gene identification when looking at potential *cis* effects within a smaller (100 kb) or larger (1 Mb) window surrounding the targeted cCRE (table S3F) (*28*–*30*).

Most *cis*-target genes were also the closest gene to the variant; however, there were 10 *cis*-target genes that were the second closest and eight that were farther away (Fig. 2D). We identified a single *cis*-target gene for 116 CREs and identified 18 CREs with two or more

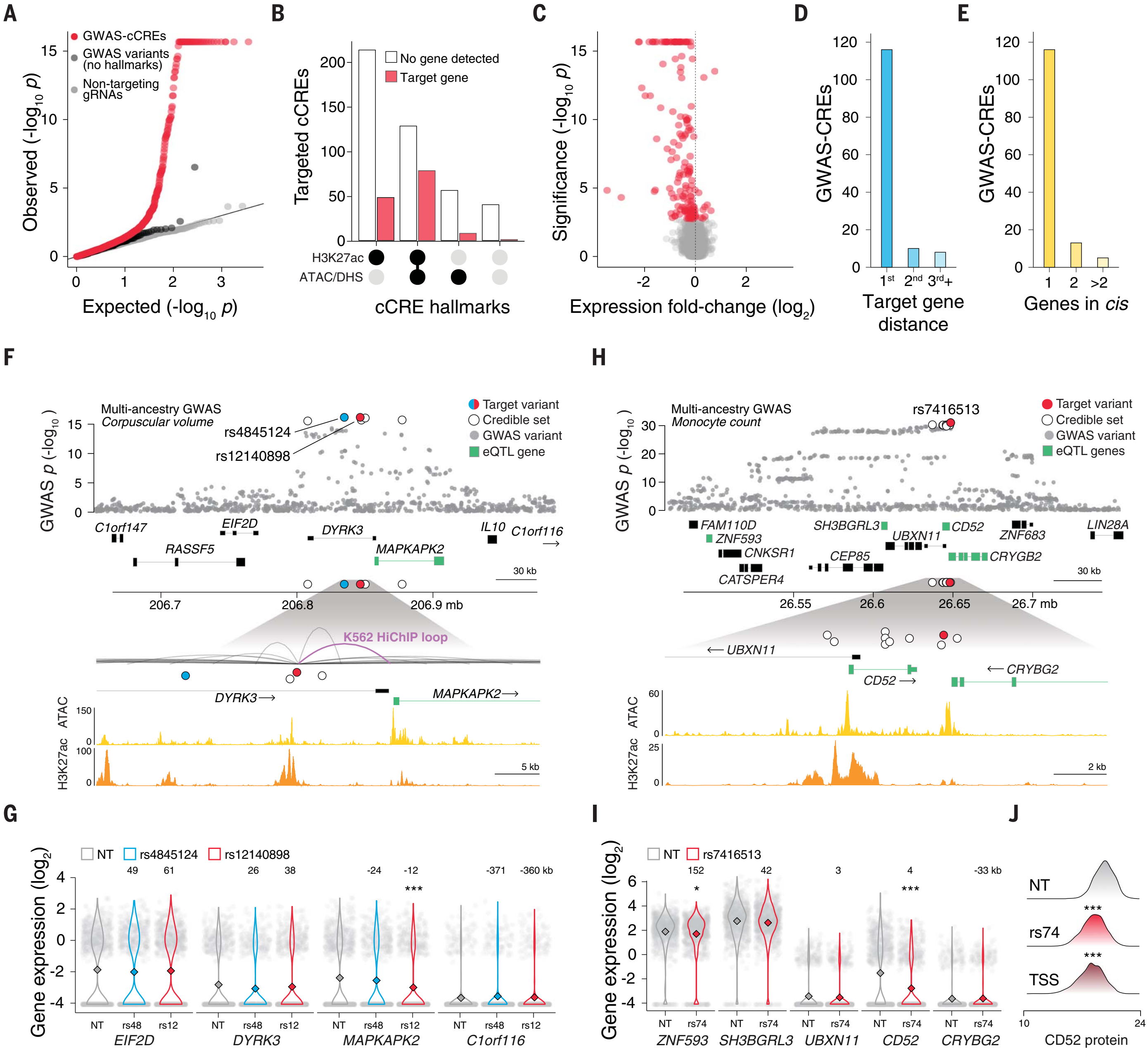


Fig. 2. Mapping *cis*-regulatory target genes for blood trait GWAS variants. (**A**) Quantile-quantile plot of *cis* effects (within 500 kb) of 531 cCREs (defined as regions with regulatory hallmarks ATAC/DHS or H3K27ac) overlapping 535 GWAS variants (GWAS-cCREs), 41 GWAS variants without CRE hallmarks, and 32 nontargeting (NT) gRNAs. Genes for NT tests were randomly sampled from the set of genes in *cis* for targeting gRNAs. We identified 154 pairs of target genes and CREs for GWAS variants with CRE hallmarks, one target gene-CRE pair for GWAS variants without CRE hallmarks, and no target genes with NT gRNAs significant at a 5% FDR (Benjamini-Hochberg–adjusted SCEPTRE *P* value). (**B**) Targeted GWAS-cCREs with and without target genes detected and their functional hallmarks of enhancer activity (ATAC/DHS or H3K27ac) in K562 cells. (**C**) Volcano plot of *cis*-regulatory effects. Significant pairs of genes and GWAS-CREs are indicated in red. (**D**) Distance to gene rank for GWAS-CREs and target genes, where genes were ranked according to closest TSS to a given GWAS-CRE. (**E**) Number of target genes in *cis* per GWAS-CRE. (**F**) Top: For a multiancestry corpuscular volume locus, two fine-mapped variants were targeted, the lead variant, rs4845124 (blue), and rs12140898 (red). *MAPKAPK2* (green) was nominated as the target gene by fine-mapped blood cell eQTLs for both variants. Bottom: rs12140898 mapped to a K562 HiChIP loop connecting its GWAS-CRE to the *MAPKAPK2* promoter. (**G**) Single-cell gene expression for cells with gRNAs targeting GWAS-cCREs (rs4845124 or rs12140898) or NT. Only rs12140898 had a target gene within 500 kb, *MAPKAPK2*. (**H**) For a multiancestry monocyte count locus, one fine-mapped variant was targeted, the lead variant, rs741613 (red). *ZNF593*, *SH3BGRL3*, *CD52*, and *CRYBG2* (green) were nominated as target genes by fine-mapped blood cell eQTLs. (**I**) Single-cell gene expression for cells with gRNAs targeting the GWAS-cCRE rs741613 or NT. *CD52* and *ZNF593* were both identified as target genes. (**J**) Single-cell protein expression for cells with gRNAs targeting the GWAS-cCRE rs741613, the *CD52* TSS, or NT. Asterisks denote significant *q* values, Benjamini-Hochberg–adjusted SCEPTRE *P* values (*$q < 0.05$, **$q < 0.01$, ***$q < 0.001$).

cis-target genes (Fig. 2E). We also targeted 41 variants that were the most plausibly causal variants at their respective loci but did not overlap biochemical hallmarks of enhancers. From the 41 variants we targeted that did not overlap called peaks for biochemical hallmarks of enhancers, there was one variant (rs106585 for WBC counts) with a significant target gene, *LTBR* [$\log_2$ fold-change (FC) = −0.38, SCEPTRE $P = 3.1 \times 10^{-7}$] (Fig. 2A, fig. S7, and table S3G). Upon further inspection, we found a weak enhancer-associated histone modification (H3K27ac) at this locus despite the lack of a called peak, suggesting that biochemical hallmarks of enhancer activity are required and that spurious signals from inactive chromatin are rare (fig. S8).

We next sought to characterize concordance between *cis*-target genes identified using STING-seq and other methods such as physical contact mapping and allele-specific expression. To identify gene promoters anchored in three-dimensional space to H3K27ac-bound chromatin, we generated H3K27ac HiChIP libraries in K562 cells. Of the 134 STING-seq CREs and their 124 target genes, we observed 32 CREs in which the same gene was identified with H3K27ac HiChIP contacts, 27 CREs in which the same gene was identified through expression quantitative trait loci (eQTL) mapping of the same fine-mapped variant (*31*), and 73 CREs in which the same gene was identified through a transcriptome-wide association study (TWAS) of a blood trait (*32*). Although the sensitivity of TWAS for target gene identification is reasonably high (54%), we and others have found that specificity can be low using this approach (*33*). Additionally, 54 CREs with fine-mapped GWAS variants had allele-specific effects on enhancer activity or TF binding (*34*, *35*), suggesting that these variants are causal at their respective CREs (table S3F).

Identification of causal variants and their impact on gene and protein expression

In the STING-seq dataset, we identified examples in which multiple lines of orthogonal evidence converged to explain how a CRE regulates a *cis*-target gene. For example, the lead variant (rs4845124) at a locus associated with mean corpuscular volume in multiancestry meta-analyses (GWAS $P = 6.9 \times 10^{-17}$) was fine-mapped as plausibly causal (in the 95% credible set with posterior probability ≥ 1%); however, upon CRISPR inhibition of the cCRE, there was no target gene (Fig. 2, F and G). Fine-mapping of this locus nominated a second plausibly causal variant mapping to a cCRE (rs12140898), and its inhibition identified *MAPKAPK2* as the target gene ($\log_2$ FC = −0.64, SCEPTRE $P = 2.2 \times 10^{-16}$). Both variants were fine-mapped eQTLs for *MAPKAPK2* in neutrophils. However, only rs12140898 had predicted allele-specific effects on SPI1 binding and mapped to a HiChIP contact domain for the *MAPKAPK2* promoter. Therefore, although eQTL studies nominated the correct target gene for this locus, it was through experimental CRE to gene mapping that we pinpointed the most likely causal GWAS variant. Most of the targeted GWAS variants did not have supporting evidence from eQTL data but were within proximity (500 kb) of a TWAS gene, demonstrating that we can uncover genes that may be underpowered by eQTL mapping and refine TWAS results that may have high false-positive rates (table S3F) (*33*).

To disentangle loci with multiple target genes in *cis*, we can combine targeted CRE inhibition and gene inhibition. For example, the lead variant (rs7416513) at a locus associated with monocyte count in multiancestry meta-analyses (GWAS $P = 3.8 \times 10^{-32}$) was fine-mapped as plausibly causal (Fig. 2H). This variant maps to an intergenic region between the gene bodies of *CRYBG2* and *CD52*, and the gene with the closest TSS is *UBXN11*. Given this, it is unclear which of these genes, if any, might be the target gene. The variant is also a fine-mapped blood cell eQTL for multiple genes in the locus (*CD52*, *CRYBG2*, *SH3BGRL3,* and *ZNF593*), further obscuring the target gene. Upon inhibiting the rs7416513-CRE, we detected *CD52* as the most significantly altered gene ($\log_2$ FC = −1.6, SCEPTRE $P = 2.2 \times 10^{-16}$) (Fig. 2I), and *ZNF593* also had a weak change in expression ($\log_2$ FC = 0.04, SCEPTRE $P = 1.3 \times 10^{-3}$), with no effect on *SH3BGRL3* or *CRYBG2*. Directly targeting *CD52* does not influence *ZNF593* (SCEPTRE $P = 0.65$) expression, suggesting the rs7416513-CRE has a pleiotropic regulatory effect on multiple genes.

Using single-cell proteomics, we also detected a significant decrease in cell surface CD52 protein expression upon rs7416513-CRE inhibition ($\log_2$ FC = −0.1, SCEPTRE $P = 1.2 \times 10^{-15}$) (Fig. 2J), demonstrating that CREs with GWAS variants modulate not only *cis*-target gene expression but also protein expression. CD52 protein can be targeted with alemtuzumab to improve clinical outcomes in patients with myelodysplastic syndrome, suggesting that this may be the causal gene for the monocyte count GWAS association (*36*). The rs7416513-derived C allele is associated with increased monocyte count in multiancestry meta-analyses (GWAS effect = 0.025, $P = 3.8 \times 10^{-32}$) (*11*) and also with increased CD52 expression in monocytes (eQTL estimate = 0.71, $P = 4.5 \times 10^{-31}$) (*37*), highlighting the power of STING-seq to connect variants to druggable genes and identify those variants that may affect response to drugs such as alemtuzumab.

Target gene discovery in STING-seq using non-European and multiancestry GWASs

Historically, most GWAS loci have been identified using individuals of European ancestry (*38*). Recent efforts to use non-European ancestries and to combine multiple ancestries for GWASs have yielded numerous new associations (*11*, *39*, *40*). By leveraging ancestry-specific and multiancestry GWASs, we increased the discovery space of CREs and target genes for STING-seq. We identified 16 CREs with *cis*-target genes from GWAS variants in non-European ancestries. For example, we identified *ATP1A1* as the target gene for a locus associated with neutrophil counts exclusively in African ancestries (fig. S9, A and B). The lead variant (rs6674304) was fine-mapped as plausibly causal in individuals with African ancestries (GWAS $P = 3.4 \times 10^{-44}$) but not in individuals with European ancestries (GWAS $P = 0.58$). Although rs6674304 did not map to any cCREs, statistical fine-mapping nominated three additional variants that did map to cCREs (rs6660743, rs12087680, and rs7544679) (fig. S9A). We targeted all three variants using STING-seq and found that targeting the rs12087680-CRE revealed the *cis*-target gene *ATP1A1* ($\log_2$ FC = −0.35, SCEPTRE $P = 2.0 \times 10^{-10}$) (fig. S9B). ATP1A1 maintains electrochemical gradients of sodium and potassium ions, and prior work has linked both *ATP1A1* and neutrophil counts with hypertension (*41*–*43*). As the *ATP1A1* CRE demonstrates, STING-seq using non-European and multiancestry GWAS can identify new trait genes.

A pleiotropic CRE in the APOE *and* APOC1 *locus*

In a minority of STING-seq CREs, we identified multiple *cis*-target genes that may occur through direct regulation of multiple genes or indirect effects on other nearby genes driven by a single *cis*-target gene. These outcomes can be difficult to distinguish without additional perturbations or a known gene-regulatory network. For example, we found that rs1065853 was the lead variant and fine-mapped as plausibly causal for an immature RBC trait (high light scatter reticulocyte percentage) at its locus (GWAS $P = 5.8 \times 10^{-48}$) (fig. S9C). This variant mapped to an intergenic region between the gene bodies of *APOE* and *APOC1*, with *APOE* being the closest gene, and was also associated with high- and low-density lipoprotein levels (*44*). Upon inhibiting the rs1065853-CRE, we observed significant decreases in expression for both *APOE* ($\log_2$ FC = −0.63, SCEPTRE $P = 2.8 \times 10^{-6}$) and *APOC1* ($\log_2$ FC = −0.27, SCEPTRE $P = 3.5 \times 10^{-6}$) (fig. S9D). Previous studies have shown that *APOE* and *APOC1*, which encode apolipoproteins E and C1, respectively, influence blood lipids and diverse ailments including cardiovascular disease and Alzheimer's disease (*45*, *46*). To help distinguish direct and indirect regulation, we used a prior genome-wide Perturb-seq (GWPS) study in the same cell line (K562) to determine whether *APOE* or *APOC1* regulate one another (*47*). *APOC1*

expression was unchanged upon *APOE* inhibition (GWPS z = 0.02), but *APOE* expression was decreased upon *APOC1* inhibition (GWPS z = −1.4). *APOE* and *APOC1* direct inhibition suggests that rs1065853-CRE may target either *APOC1* alone (even though *APOE* is the closest gene) or both *APOC1* and *APOE*. Because these genes work in a coordinated fashion to regulate lipid metabolism (48), the co-regulation of these genes is a notable observation of regulatory pleiotropy that may contribute to trait associations.

Targeting multiple CREs in the PTPRC *locus reveals nonfunctional linkage disequilibrium proxies*

We also examined loci with several fine-mapped variants near a single gene. At the *PTPRC* locus, we targeted nine variants that were fine-mapped variants for 10 traits (fig. S10A and table S1E) and mostly not in strong linkage disequilibrium (LD), as quantified by pairwise R^2 from 1000 Genomes (49) (fig. S10B). The nine variants mapped to distinct cCREs: One was 5 kb before the *PTPRC* TSS and the remaining eight were in the first intron, from 2 to 42 kb after the TSS (fig. S10C). We observed modulation in *PTPRC* when targeting six of the cCREs (fig. S10D). For the cCREs with no effect, we found that two variants were in high LD ($R^2 \geq 0.95$) with variants mapping to *PTPRC* CREs, suggesting that these may be nonfunctional variants in LD with functional variants (i.e., nonfunctional LD proxies). For all CREs, *PTPRC* was the only significant target gene and thus is very likely the causal GWAS gene (table S1E).

The high allelic heterogeneity, which is driven by multiple independent regulatory variants in distinct CREs modulating *PTPRC* expression, and the 10 blood trait associations suggest that the CREs may have cell type–specific activity. That is, different CREs may regulate *PTPRC* in different contexts, given that the 10 trait associations include RBCs, WBCs, and platelet traits (fig. S10A).

We found that experimental evidence (e.g., STING-seq) is required to link these CREs to *PTPRC* expression. None of the targeted variants were fine-mapped blood eQTLs, and only a single targeted variant, rs1326279, showed evidence of allele-specific effects on TF binding (31, 35). Thus, in silico methods that use eQTL data are insufficient to measure the impact of the CREs on *PTPRC* expression.

Direct GWAS variant insertion with beeSTING-seq

Next, we sought to expand the STING-seq approach to precise insertion of fine-mapped GWAS variants with base editing. We fused a cytosine base editor (FNLS-BE3) to a protospacer-adjacent motif (PAM)–flexible Cas9 variant (SpRY) (table S1F) and validated activity using gRNAs designed to disrupt splice junctions in *CD46*, which encodes a ubiquitously expressed cell surface protein, in an arrayed fashion (Fig. 3, A and B, and table S3H) (50, 51). We observed up to ~70% knockdown of CD46 when targeting splice sites with diverse PAM sequences and an average knockdown of 27% (n = 12 target sites), similar to prior pooled screens using base editing (52, 53), (fig. S11 and table S3H). We then performed a single-cell pooled base editing screen (beeSTING-seq) targeting 46 C>T fine-mapped GWAS variants mapping to 42 STING-seq–identified CREs with three gRNAs each (table S3I). We tested for direct effects on known target genes and found that 32 of 46 had at least two gRNAs with concordant effects and that all three gRNAs had concordant effects for 17 variants (Fig. 3C and table S3, J and K). We identified three sets of beeSTING-seq gRNAs with *cis*-regulatory effects on the same target genes identified using STING-seq (5% FDR) with no enrichment of nontargeting (negative control) gRNAs (Fig. 3D and table S3, L and M).

In one case, beeSTING-seq gRNAs targeted the lead variant (rs142122062) at a locus associated with RBC volume in multiancestry meta-analyses (GWAS $P = 8.2 \times 10^{-11}$) (Fig. 3E and table S3M). Targeted inhibition of the rs142122062-CRE decreased *APPBP2* expression ($\log_2$ FC = −0.46, SCEPTRE $P = 2.5 \times 10^{-4}$) and identified it as the target gene for this locus (Fig. 3F). For beeSTING-seq, we were able to design multiple gRNAs capable of inserting the same single-nucleotide edit by capitalizing on the targeting flexibility of SpRY Cas9 (51). With direct insertion of the rs142122062-T allele with two independent gRNAs, we observed a significant increase in *APPBP2* expression (combined $\log_2$ FC = 0.74, SCEPTRE $P = 7.6 \times 10^{-5}$) (Fig. 3G), demonstrating the ability of beeSTING-seq to identify GWAS variants that act to increase expression. Both gRNAs exclusively edit the GWAS variant, because it is the only C nucleotide within the editing window (50). Using TWAS, we found that amyloid precursor protein, which APPBP2 binds, had the strongest association with RBC counts (54), suggesting a possible mechanism of how altered *APPBP2* expression affects RBC traits. In this manner, beeSTING-seq can more precisely determine the effect of GWAS variants, moving beyond CRE inhibition to reveal the impact of specific alleles on target gene expression.

CRE-driven, dosage-dependent, transcriptome-wide changes in gene expression

To understand the impact of GWAS-CREs on gene expression across the genome, we performed transcriptome-wide differential expression tests. We applied a strict (1%) FDR to identify target genes in *trans* and again found good calibration with nontargeting gRNAs (Fig. 4A and table S3C). We observed *trans* effects for CREs that targeted in *cis* the TFs *GFI1B*, *NFE2*, *IKZF1*, *HHEX*, and *RUNX1* and the host genes of microRNAs (miRNAs) *miR-142* and *miR-144/451* (Fig. 4A and tables S3F and S4A). These TFs and miRNAs are known to play key roles in hematopoietic stem cell differentiation (55–61).

For *GFI1B*, we identified two independent CREs with *trans* effects. One variant (rs524137), associated with monocyte percentage and basophil counts, maps to an intergenic CRE 11.5 kb downstream of *GFI1B* (Fig. 4B). The other variant (rs73660574), associated with several RBC traits (mean sphered corpuscular volume, immature reticulocyte fraction, mean reticulocyte volume, and mean corpuscular hemoglobin), maps to a CRE in an intron of *GFI1B* (Fig. 4B). These CREs exhibited independent dosage effects on *GFI1B* expression, with the rs524137-CRE having an ~70% stronger effect than the rs73660574-CRE. Thus, perturbing either rs73660574- or rs524137-CREs led to changes in the expression of *GFI1B* (Fig. 4C) and its target genes. To better understand the *trans* effects of these two *GFI1B* CREs, we examined gene-expression changes in all 1161 differentially expressed genes identified from the rs524137-CRE (Fig. 4D). For these genes, we observed a high correlation between perturbations targeting each CRE (r = 0.84), even though many of the gene expression changes were more modest when perturbing the rs73660574-CRE. We found a linear dosage relationship between the *trans*-regulatory effects for the CREs that agreed with the difference in their effect on *cis* (*GFI1B*) expression (~1.3-fold) (Fig. 4, C and D). Using single-cell proteomics in the same cells, we observed changes in protein levels for nine of the genes in the *GFI1B* network; for these, changes in transcript expression and protein levels were highly correlated (r = 0.9) (fig. S12). This example demonstrates how GWAS variants mapping to CREs perturb regulatory networks, and that these changes at the RNA level also alter protein expression.

In addition to *GFI1B*, we also observed CRE dosage effects on target gene expression and regulatory networks for *NFE2* (rs79755767, associated with hematocrit and red cell distribution width, and rs35979828, associated with eosinophil count, mean corpuscular hemoglobin, and monocyte count) (Fig. 4E). When targeting these variants, we observed dosage effects on *NFE2* expression (rs79755767-CRE $\log_2$ FC = −1.1, SCEPTRE $P = 2.2 \times 10^{-16}$; rs35979828-CRE $\log_2$ FC = −0.6, SCEPTRE $P = 2.2 \times 10^{-16}$) (Fig. 4F) and on a 343-gene regulatory network (r = 0.78) (Fig. 4G). These results reinforce our findings that fine-mapped GWAS variants at independent CREs have independent effects, not only on target gene

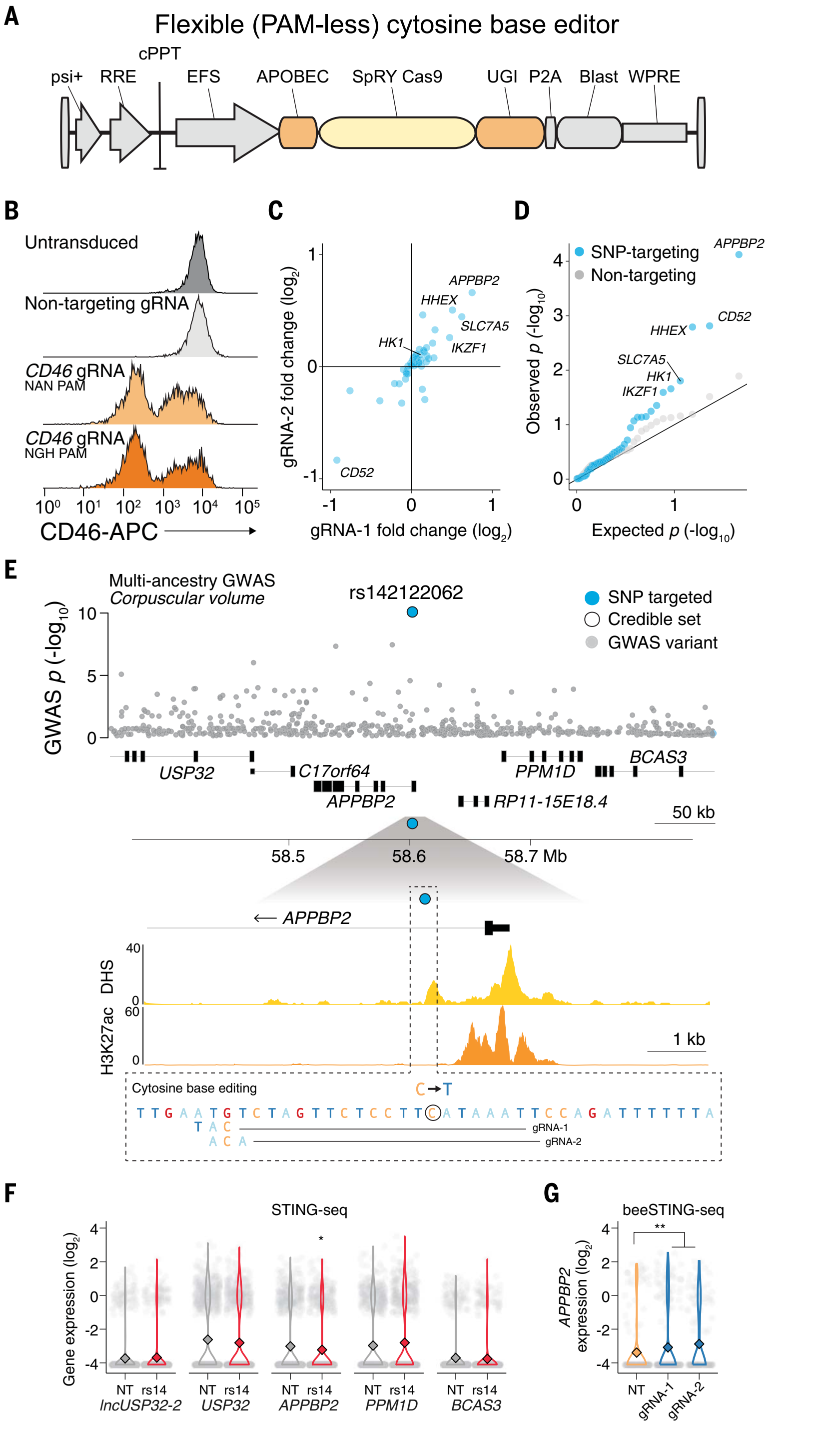


Fig. 3. Precise GWAS variant editing with beeSTING-seq. (**A**) Lentiviral CRISPR base editor (FNLS-BE3) with a relaxed PAM SpCas9 variant, SpRY, for beeSTING-seq. (**B**) Flow cytometry of CD46 cell surface protein after base editing at *CD46* splice donor sites. CD46 knockdown was compared with untransduced and NT controls. (**C**) Target gene fold change for the two gRNAs with the most concordant effects for each variant. (**D**) Quantile-quantile plot of NT gRNAs and gRNAs targeting 46 fine-mapped GWAS variants mapping to STING-seq GWAS-CREs with *cis*-effect genes. Genes for NT tests were randomly sampled from the set of genes in *cis* for targeting gRNAs. (**E**) Top: For a multiancestry corpuscular volume locus, one fine-mapped variant was targeted, the lead variant, rs142122062 (blue). Bottom: Base editing by gRNA-1 and gRNA-2 changes the rs142122062 allele from reference to alternative; for both gRNAs, this is the only cytosine in the base editing window. (**F**) Single-cell gene expression for cells with gRNAs targeting the GWAS-cCRE rs142122062 or NT. *APPBP2* was identified as a *cis*-target gene. (**G**) beeSTING-seq of rs142122062 increases *APPBP2* expression with two independent gRNAs with positions shown in (E). Asterisks denote significant *q* values, Benjamini-Hochberg–adjusted SCEPTRE *P* values (**q* < 0.05, ***q* < 0.01, ****q* < 0.001).

expression, but also on entire regulatory networks in *trans*.

A limitation of many GWAS functional interpretation approaches is that they focus on nearby protein-coding genes and overlook relevant noncoding RNAs. With STING-seq, we also identified regulatory networks for miRNAs, which can have a broad impact on gene regulation. For example, STING-seq at the CRE for rs2526377, the most plausibly causal variant for a locus associated with platelet count locus, revealed no protein-coding *cis*-target genes (fig. S13A). However, when examining noncoding transcripts, we found a differentially expressed noncoding transcript, *AC004687.1*, which is also known as the *miR-142* host gene ($\log_2$ FC = –1.8, SCEPTRE $P = 2.2 \times 10^{-16}$) (fig. S13B). This finding is further supported by prior work in the context of Alzheimer's disease showing that the risk allele decreases *miR-142* host gene promoter activity (*62*, *63*).

For STING-seq perturbation of rs2526377, we detected a 119 gene *trans*-regulatory network (fig. S13C). The top up-regulated genes within the rs2526377 *trans*-regulatory network (*WASL* and *CFL2*) were also the top up-regulated genes in miR-142 knockout mice (*60*). This lends further support that the *trans*-regulatory effects of rs2526377 perturbation are caused by *cis* effects on miR-142, as found in STING-seq. This *cis*-target miRNA and its regulatory network can be easily missed when considering only protein-coding genes for target gene annotation.

We also analyzed *trans* effects with direct variant insertion using beeSTING-seq. We could detect changes in regulatory network expression in the expected direction upon inserting the rs12784232-A allele (associated with lymphocyte percentage) and the rs6592965-A allele (corpuscular hemoglobin), which mapped to the *HHEX* and *IKZF1* GWAS-CREs, respectively. In contrast to GWAS-CRE inhibition, which decreased expression of *HHEX* and *IKZF1* (Fig. 4A), direct variant insertion resulted in increased expression of the *cis*-target genes and, accordingly, *trans* effects for genes tended to switch directions in differential expression compared with STING-seq. Specifically, we observed that 60 to 70% of *HHEX* and *IKZF1* network genes had reversed directions of effect, demonstrating that GWAS variants that act to increase expression can affect networks in discordant directions from CRE silencing.

Enrichment of cis*-target binding sites and GWAS genes in* trans*-regulatory networks*

To better characterize how CREs with target genes in *trans* alter blood cell phenotypes, we examined genome-wide binding for GFI1B, NFE2, IKZF1, and RUNX1 (ChIP-seq) (*64*, *65*) and sequence-based predicted targets of miR-142 and miR-144/451 (TargetScan) (*66*, *67*).

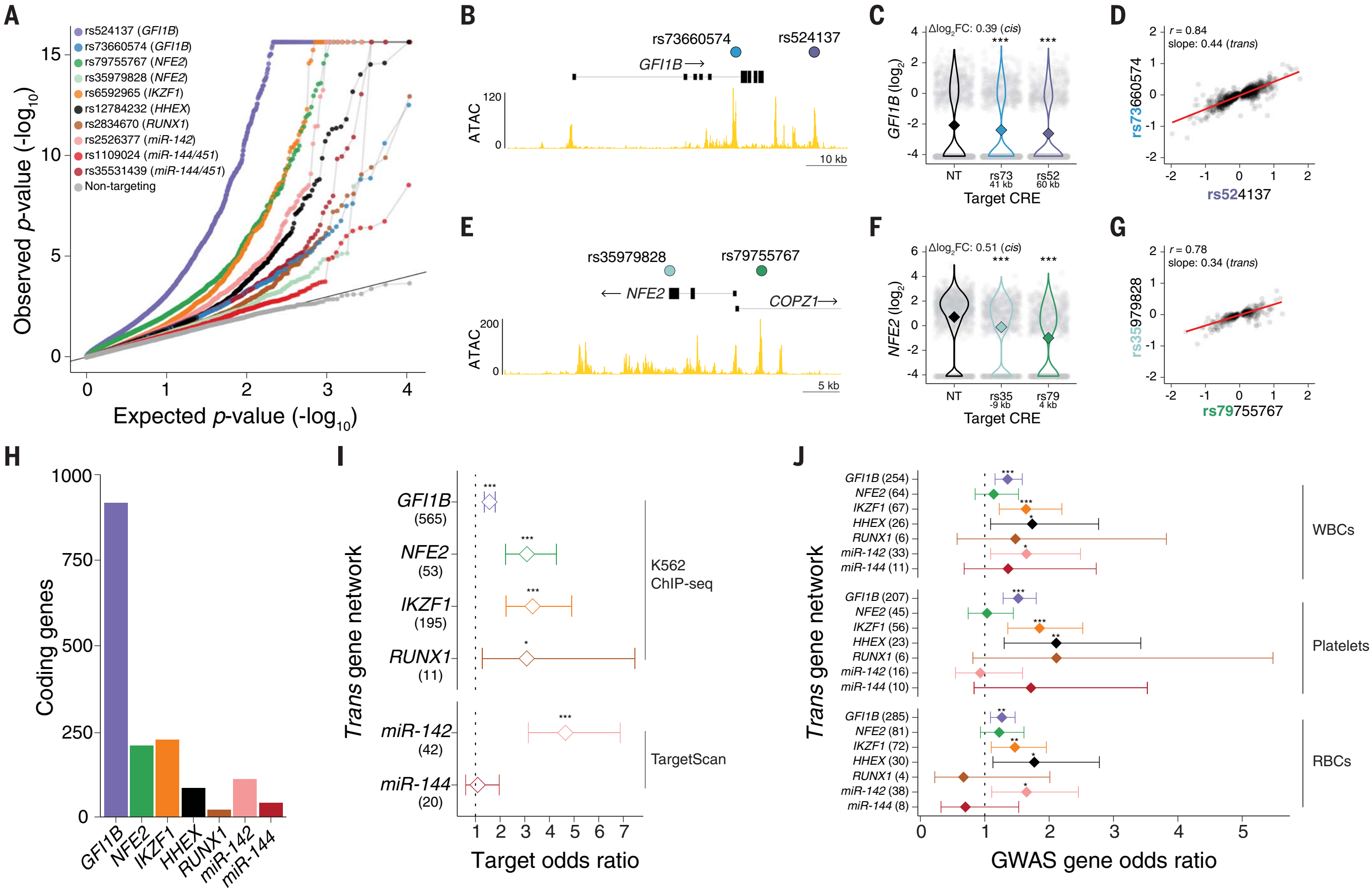


Fig. 4. *Trans*-regulatory network discovery of genes that affect diverse blood cell traits. (**A**) Quantile-quantile plots of *trans* effects (whole transcriptome) of GWAS-CREs and NT gRNAs. We identified significant genes at a 1% FDR (Benjamini-Hochberg–adjusted SCEPTRE *P* value) for GWAS-CREs with TFs (*GFI1B*, *NFE2*, *IKZF1*, *HHEX*, and *RUNX1*) and miRNAs (*miR-142* and *miR-144/451*) as *cis*-regulatory target genes. No *trans* effects were found for NT gRNAs. (**B**) Two GWAS-CREs targeted at the *GFI1B* locus, rs524137 and rs79755767. (**C**) Single-cell gene expression for cells with gRNAs targeting GWAS-cCREs at the *GFI1B* locus (rs524137 or rs79755767) or NT. (**D**) Expression of rs524137-CRE significant *trans*-target genes in cells with perturbation of either GWAS-CRE at the *GFI1B* locus (rs524137 or rs79755767) (n = 1161 genes at a 1% FDR). (**E**) Two GWAS-CREs targeted at the *NFE2* locus, rs79755767 and rs35979828. (**F**) Single-cell gene expression for cells with gRNAs targeting GWAS-cCREs at the *NFE2* locus (rs79755767 or rs35979828) or NT. (**G**) Expression of rs79755767-CRE significant *trans*-target genes in cells with perturbation of either GWAS-CRE at the *GFI1B* locus (rs79755767 or rs35979828) (n = 343 genes at a 1% FDR). (**H**) Protein-coding genes with changes in expression for *trans*-regulatory networks. (**I**) Gene set enrichment odds ratios (diamonds) and 95% confidence intervals (lines) for TF and miRNA targets within each *trans*-regulatory network. Targets for TFs were given as the closest gene to each TF-specific ENCODE K562 ChIP-seq peak and for miRNAs as TargetScan-predicted targets on the basis of sequence. (**J**) For each *trans*-regulatory network, gene set enrichment odds ratios (diamonds) and 95% confidence intervals (lines) of closest genes to fine-mapped variants from WBC, platelet, and RBC GWASs from 29 UKBB blood traits and 15 BCX blood traits. Asterisks: (C) and (F), significant Benjamini-Hochberg–adjusted SCEPTRE *P* values; (I) and (J), logistic regression *P* values (**P* < 0.05, ***P* < 0.01, ****P* < 0.001).

We investigated whether the closest genes to each ChIP-seq peak or predicted miRNA target genes were enriched in STING-seq *trans*-regulatory networks (Fig. 4H and table S4B). We observed enrichments of predicted target genes for *GFI1B*, *NFE2*, *IKZF1*, *RUNX1*, and *miR-142* (odds ratio = 2.4 ± 1.9, mean ± SEM) (Fig. 4I and table S4C). Thus, perturbing CREs can reveal second-order interactions for regulatory networks driven by TFs or miRNAs.

A related and pertinent question is whether the genes in the *trans*-regulatory networks identified by STING-seq may also play a role in blood traits and if they also harbor *cis*-regulatory genetic variants. To answer this question, we constructed a set of putatively causal genes for each of the 29 UKBB and 15 BCX GWASs by selecting the closest protein-coding genes to fine-mapped variants of GWAS loci. We then grouped them by cell type, generating gene sets for platelets, RBCs, and WBCs that were mostly distinct (fig. S14 and table S4B). For nearly all *trans*-regulatory networks, we found enrichments for blood cell GWAS genes (Fig. 4J and table S4C). These blood cell trait GWAS loci enrichments indicate that the known roles of these genes in hematopoiesis and cell differentiation are mediated by their effects on regulatory networks. Furthermore, identification of the *trans* genes with STING-seq pinpointed regulatory networks for which polygenic perturbation by distinct variants across the genome appears to contribute to the GWAS signal. This suggests a mechanistic importance for networks themselves, for which we do not need to functionally determine V2F per locus if we know the pathway through which they are likely to act, similar to recent work focusing on perturbation of target genes (*68*).

Trans-regulated genes reveal biological mechanisms and cell types of trait associations

Given these relationships between *trans*-regulated genes and GWAS loci, we analyzed the structures of these regulatory networks to better

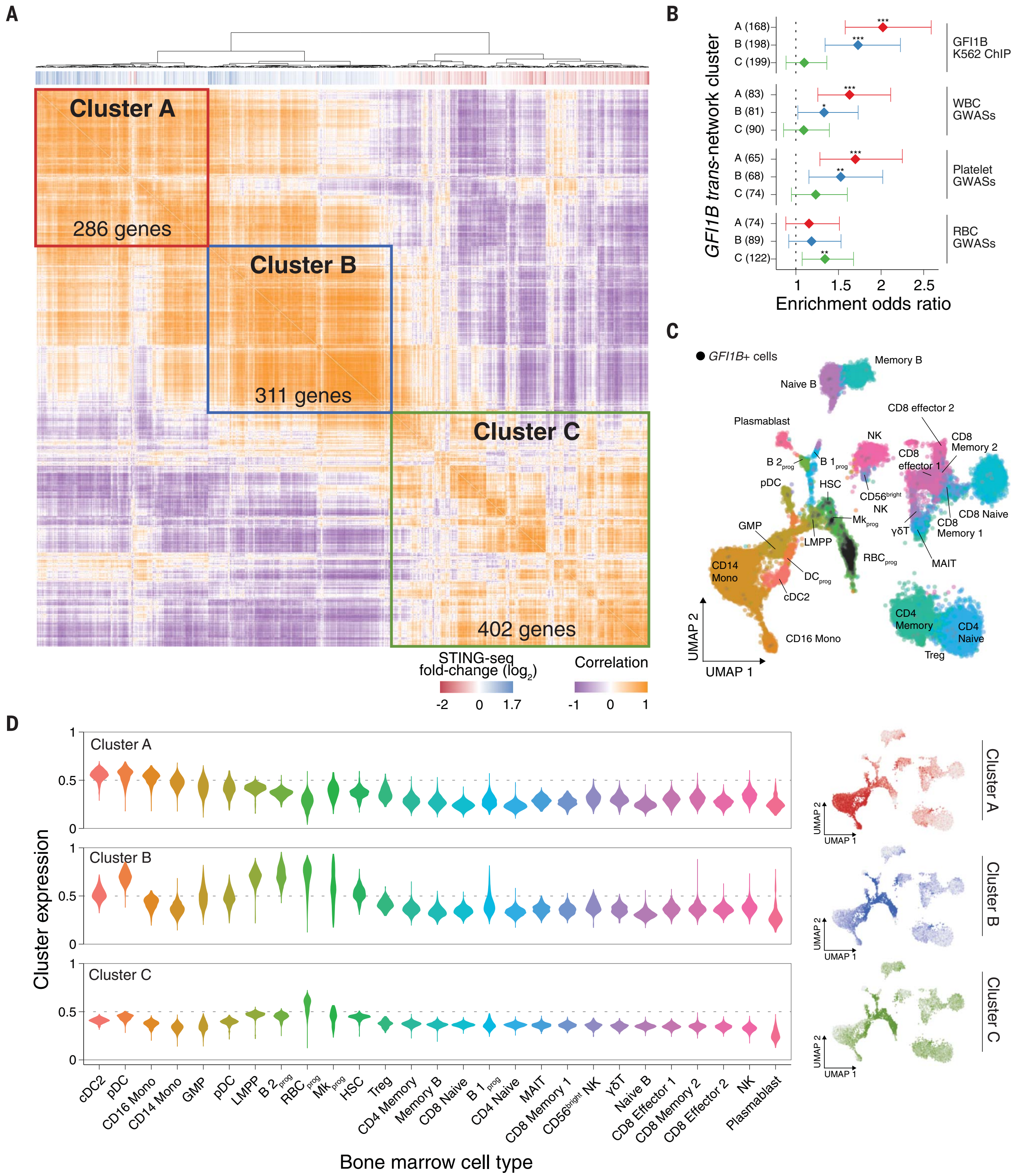


Fig. 5. Subnetworks of *GFI1B* target genes are expressed in specific hematopoietic progenitors and differentiated cells. (**A**) Coexpression matrix of rs524137-CRE *GFI1B* network genes in K562 with hierarchical clustering. Three clusters (A, B, and C) are indicated. The vertical bars below the dendrogram indicate if genes had increased (blue) or decreased (red) expression upon inhibiting the *GFI1B* CRE. (**B**) For each *trans*-regulatory *GFI1B* subnetwork (cluster), gene set enrichment odds ratios (diamonds) and 95% confidence intervals (lines) of closest genes to GFI1B K562 ChIP-seq peaks (top) and fine-mapped variants (bottom) from WBC, platelet, and RBC GWASs from 29 UKBB blood traits and 15 BCX blood traits. Asterisks denote logistic regression *P* values (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$). (**C**) UMAP of human bone marrow cell gene expression from 35 Human Cell Atlas donors. Labels and colors indicate cell types. B 2_{prog}, progenitor B-2 cells; RBC_{prog}, RBC progenitors; DC_{prog}, dendritic cell progenitors (for full list, see table S4D). The black dots denote cells expressing *GFI1B*. *GFI1B* is most highly expressed in RBC progenitors, megakaryocyte progenitors, and hematopoietic stem cells. (**D**) Expression of genes from clusters A, B, and C in each human bone marrow cell type (left) and in each cell in the UMAP space from (C) (right).

understand the mechanistic roles of specific genes in blood traits. Using single-cell gene coexpression and clustering, we identified coexpressed gene clusters for each of the loci (Fig. 5A and fig. S15). For the *trans*-acting gene *GFI1B*, we identified two clusters (A and B) of genes with increased expression upon *GFI1B* CRE repression with STING-seq. These clusters were the most strongly enriched for *GFI1B* binding sites (Fig. 5B and table S4, B and C). A third cluster (C) consisted primarily of genes with decreased expression, which were not enriched for GFI1B-binding sites. Clusters A and B were enriched for genes from platelet and WBC GWASs, whereas cluster C was only enriched for genes from RBC GWASs.

To further refine and validate the individual cell types involved with different clusters of coregulated genes, we integrated the *GFI1B* coexpression network with primary cells from the Human Cell Atlas, which includes progenitors and/or differentiated cell types for platelets, WBCs, and RBCs. Specifically, we used single-cell RNA sequencing from 35 bone marrow donors (*69*, *70*), because bone marrow includes a rich sample of multipotent progenitor cells crucial for hematopoiesis. We first confirmed that *GFI1B* was expressed in hematopoietic stem cells and progenitor cells for RBCs and megakaryocytes, consistent with *GFI1B*'s well-established role as a transcriptional repressor in early and lineage-specific progenitors (Fig. 5C) (*55*, *71*–*73*). As expected, *GFI1B* was not expressed in granulocytes and lymphocytes (*73*, *74*). Genes from cluster A were highly enriched for GFI1B-binding sites and had increased expression upon inhibiting *GFI1B*, suggesting that these genes are actively repressed in cells in which *GFI1B* is expressed (Fig. 5B). We next observed that genes from cluster A were highly expressed in granulocyte-monocyte progenitors and differentiated WBC types, including monocytes and dendritic cells (Fig. 5D and table S4D). For example, *CD33* is a well-known marker for myeloid cells that is commonly used to diagnose acute myeloid leukemia, and its expression increases upon inhibiting the *GFI1B* CRE (fig. S12) (*75*, *76*). GFI1B directly binds the promoter of *CD33* (fig. S16A) and, upon inhibiting *GFI1B*, we found that *CD33* transcript and protein expression were both increased (fig. S16, B and C). *CD33* is expressed in myeloid progenitors and differentiated cells such as dendritic cells or monocytes (fig. S16D). Overall, cluster A is composed of genes that *GFI1B* directly represses, along with their downstream targets, to prevent differentiation of hematopoietic stem cells into WBCs.

Like cluster A, genes in cluster B were also enriched for GFI1B-binding sites and had increased expression upon inhibiting *GFI1B* (Fig. 5, A and B). However, genes in cluster B were not expressed in differentiated WBCs, but rather in a broad set of progenitor cell types (Fig. 5D), suggesting that these may be genes that are repressed in hematopoietic stem cells to maintain a multipotent cell state. Cluster C differed from clusters A and B in that it was not enriched for GFI1B-binding sites and had decreased expression upon inhibiting *GFI1B*. Genes in cluster C were expressed most highly in RBC progenitors, suggesting that these genes are secondary targets of *GFI1B* that act in a lineage-specific manner to differentiate hematopoietic stem cells into erythrocytes. These findings are supported by this cluster being enriched for RBC GWAS genes (Fig. 5B), and pathway analysis identifying these genes as part of the heme biosynthesis pathway (table S4E). The identification of these *trans*-regulatory networks in a homogeneous blood progenitor–like cell type (K562) demonstrates the utility of STING-seq in studying diverse effects of CREs on target genes.

Trade-offs between CRE effect sizes, number of cells, and sequencing depth in STING-seq

Given the large number of GWASs performed over the past 15 years, with numbers of trait-associated loci per GWAS ranging from tens to thousands (*44*), we wanted to understand the scale of cells needed to perform STING-seq under various settings. By performing statistical down-sampling experiments on the *cis*-regulatory effects identified by STING-seq, we computed the number of cells required for nominal significance (SCEPTRE $P < 10^{-3}$) for target genes with different expression levels, different CRE perturbation effect sizes, and different per-cell sequencing depths (fig. S17). For CREs with large effects, STING-seq requires as few as 100 cells and 5000 reads per cell, comparable to methods such as Perturb-seq and ECCITE-seq which target genes directly (*47*, *68*, *77*, *78*). For CREs with moderate effects, STING-seq requires about 400 cells per gRNA or, if the cell number is fixed at 100 cells, 15,000 reads per cell. This down-sampling analysis provides a useful set of guidelines for estimating the resources required for applying STING-seq to other GWASs beyond blood traits.

Discussion

We have developed an approach for the characterization of functional effects of GWAS loci that takes noncoding human genetic variants and integrates fine-mapping, pooled CRISPR screens and single-cell RNA and protein sequencing to identify target genes in *cis* and *trans*. We have demonstrated the utility of STING-seq to identify target genes of CREs overlapping GWAS variants and described complex regulatory architectures of CREs. We found that 77% of blood trait GWAS loci have at least one fine-mapped variant overlapping an enhancer region and can be targeted with STING-seq. We identified target genes for 25% of tested cCREs and 36% of tested loci, a high yield over previous studies on the regulatory effects of noncoding genomic loci (*7*, *8*). We also found that CRE activity is needed for CRISPRi-based target detection, and that spurious signals from inactive chromatin are rare. Additionally, we identified CREs with GWAS variants for TFs and miRNAs and, through their perturbation, identified *trans*-regulatory network clusters with distinct biological functions. The enrichment of genes in independent blood cell trait GWAS loci in these networks implies a polygenic contribution to the cellular functions that underlie diverse blood cell traits. We also identified target genes for non-European associations for which functional genomics data are typically sparse. For example, we nominated *ATP1A1* as a causal gene for neutrophil counts by targeting a locus identified exclusively in African ancestries. Targeting loci identified from ancestry-specific GWASs in cell models is ancestry agnostic if the GWAS variant maps to a candidate regulatory element and can lead to target gene identification.

We also performed direct variant insertion with beeSTING-seq, identifying noncoding GWAS variants with causal effects on target gene expression. Given the incomplete editing efficiencies [with many studies reporting ~30% (*79*)], the fact that the biological effect of individual GWAS variants is expected to be small, and that single-cell transcriptome data are sparse, it was not unexpected that we were only able to identify few loci, and future work is needed to further optimize base editors for studying the effects of GWAS variants. Targeted enrichment panels will have utility in improving the sparsity of single-cell sequencing; however, further innovation will be necessary to improve base-editing efficiency through directed evolution of existing base editors and the discovery of additional ones. However, the trade-off between higher yield from blunt perturbations such as CRISPRi versus highly precise base editing with smaller functional effects is likely to persist, and the ideal approach depends on the goals and design of each study.

A key feature of recent CRISPRi screens of cCREs (*7*, *8*), including STING-seq, is the introduction of multiple perturbations per cell. This substantially increases the number of loci that can be feasibly analyzed. Although this is feasible for immortalized cell lines, expanding multiple perturbations (using either high MOI transduction or innovative vector designs) to other cell lines and primary cells will be instrumental for the next stage of target gene identification and characterization for diverse GWAS traits. However, caution is warranted in study

designs in which a large proportion of gRNAs are likely to have *trans* effects, because their potential interactions may complicate interpretation of the data. In these cases, reducing the number of perturbations per cell may be necessary.

Our results demonstrate the power of single-cell sequencing for sensitive and scalable readout of regulatory effects of GWAS loci in *cis* and *trans*. Although we have a high yield in *cis* target gene discovery, identification of a *cis* gene alone with STING-seq does not prove its mechanistic causal role driving the GWAS association, nor does it exclude other potential causal variants, CREs, and genes, including in other cell types. Indeed, our observation of multiple CREs with highly correlated *cis* and *trans* effects but GWAS associations for different blood traits suggests that they might have distinct additional effects in other cellular contexts. In loci in which *cis* effects are coupled with *trans*-network effects, STING-seq can be highly informative of potential cellular mechanisms, which also provides strong support for the causal role of the *cis*-target gene. Given these network enrichments, we suggest that GWAS loci that putatively target TFs or miRNAs should be high-priority targets for STING-seq given the wealth of information that we could gain. Furthermore, integration of STING-seq with cellular phenotype screens (*80*–*82*) will be an invaluable next step to connecting genetic variants with cellular mechanisms driving GWAS associations.

The STING-seq workflow provides a roadmap to addressing V2F challenges and identifying target genes for GWAS loci in a high-throughput fashion, enabling a deeper understanding of human noncoding genome function and translation of these insights into new therapies.

Materials and Methods

UKBB genome-wide association studies of blood cell traits

UKBB data were used upon ethical approval from the Northwest Multi-Centre Research Ethics Committee, and informed consent was obtained from all participants before participation. We used GWAS summary statistics for 29 blood cell traits from 361,194 white British UKBB participants: WBC (leukocyte) count, RBC (erythrocyte) count, hemoglobin concentration, hematocrit percentage, mean corpuscular volume, mean corpuscular hemoglobin, mean corpuscular hemoglobin concentration, RBC (erythrocyte) distribution width, platelet count, platelet crit, mean platelet (thrombocyte) volume, platelet distribution width, lymphocyte count, monocyte count, neutrophil count, eosinophil count, basophil count, lymphocyte percentage, monocyte percentage, neutrophil percentage, eosinophil percentage, basophil percentage, reticulocyte percentage, reticulocyte count, mean reticulocyte volume, mean sphered cell volume, immature reticulocyte fraction, high light scatter reticulocyte percentage, and high light scatter reticulocyte count (table S1A). Each GWAS was performed by fitting the following covariates to inverse normal transformed traits with linear regression models: Principal components 1 through 20, sex, age, age^2, sex and age interaction, and sex and age^2 interaction. The summary statistics were generated by the Neale Lab (www.nealelab.is/uk-biobank).

Statistical fine-mapping of UKBB blood cell traits

The 29 UKBB GWASs of blood cell traits were uniformly processed with a statistical fine-mapping pipeline. First, each GWAS was analyzed with GCTA-COJO v.1.93.1 (*13*, *14*) to identify conditionally independent lead variants (COJO $P < 6.6 \times 10^{-9}$) and define 1-Mb regions for statistical fine-mapping. All variants within 500 kb of a lead variant were analyzed with FINEMAP v.1.3.1 (*83*), a Bayesian fine-mapping method that assigns each variant a Bayes factor for being plausibly causal. Both GCTA-COJO and FINEMAP require population-matched covariance matrices, so we generated these with PLINK v.2.0 (*84*), QCTOOL v.2.0.2, BGENIX v.1.1.5 (*85*), and LDstore v.1.1 (*86*) using a subset of 50,000 UKBB white UK participants (UKBB accession code 47976). FINEMAP allows for a maximum number of causal configurations to test for each input set of variants, so we set the maximum to 10 causal configuration variants per fine-mapped region and excluded cases for which FINEMAP failed to converge. We then retained noncoding variants with a high Bayes factor ($\log_{10}$ BF ≥ 2) and that were at least 1% likely to be causal for a set of causal variants. Fine-mapped variants that had more than one Bayes factor because they were within 500 kb of multiple lead variants had their highest value retained. Across all 29 GWASs, we identified 827 loci, separated by at least 500 kb, and 57,531 fine-mapped variants. The Variant Effect Predictor (VEP) tool (*87*) was used to identify 53,874 noncoding variants.

Fine-mapped BCX blood cell trait GWAS

The BCX generated GWAS summary statistics and fine-mapped 95% credible sets for 15 blood traits from 746,667 participants from five global populations (European ancestries, South Asian ancestries, Hispanic ancestries, East Asian ancestries, and African ancestries): RBC count, hemoglobin, hematocrit, mean corpuscular volume, mean corpuscular hemoglobin, MCH concentration, RBC distribution width, WBC count, neutrophils, monocytes, lymphocytes, basophils, eosinophils, platelet count, and mean platelet volume (*11*). Each GWAS was performed within each global population by fitting linear mixed models, adjusting for cohort-specific covariates, to generate population-specific GWAS summary statistics. Population-specific GWAS were fine-mapped using an approximate Bayesian approach (*88*) to construct 95% credible sets from all variants within 250 kb of a lead variant. The 95% credible sets were generated by ordering marginal variant posterior probabilities from highest to lowest and retaining variants until the probabilities summed 95%. Population-specific GWASs for each trait were then meta-analyzed using a multiancestry meta-analysis method (*89*) that also generates marginal variant posterior probabilities, from which multiancestry 95% credible sets were generated. We additionally required that variants were at least 1% likely to be causal. Across all 15 multiancestry meta-analyzed GWASs, we identified 1191 loci, separated by at least 500 kb, and 62,494 fine-mapped variants. VEP (*87*) was used to identify 58,573 noncoding variants.

Functional annotation of causal noncoding single-nucleotide polymorphisms

We integrated multiple functional genomics datasets for K562 cells. Specifically, we used DNase I–hypersensitive sites (DHS) from ENCODE (*65*), H3K27ac ChIP-seq peak calls from ENCODE, and ATAC-seq peak calls that we generated previously (*81*) to identify candidate cCREs. We used bedtools v.2.25.0 (*90*) and bedops v.2.4.3 (*91*) to identify variants mapping directly to cCREs. We also required variants to be farther than 1 kb from any gene TSS. We analyzed the UKBB and BCX GWAS variants separately. For UKBB GWASs, we identified 10,628 distinct variants mapping cCREs in 629 loci. We then selected 88 variants from 56 loci for targeting on the basis of whether a variant was targetable and more plausibly causal than others for a given GWAS and locus by ranking FINEMAP $\log_{10}$ Bayes factors and manual inspection of loci. For the 88 selected variants, 32 were the most probable variant for at least one GWAS locus, and 52 were in the top-10 most probable variants. For the 56 loci, there was a median of 10.5 (± 8.6) targetable single-nucleotide polymorphisms (SNPs). Elements of manual inspection included selecting variants that mapped to intergenic regions between gene TSSs or selecting multiple variants that map proximal to the same gene. For BCX GWASs, we identified 10,446 variants mapping to 886 loci. We selected 507 variants mapping to 265 loci for targeting, including 41 variants mapping to closed chromatin. Of the cCRE-mapping variants, we targeted 137 that were the sole variant within the 95% credible set and 239 variants that were composed of all targetable 95% credible set variants for 112 loci. The remaining 131 variants were selected because they were identified by GWASs from non-European ancestries and either fine-mapped in a population-specific GWAS or in the multiancestry meta-analysis. K562 DHS peaks and H3K27ac, RUNX1, IKZF1,

and NFE2 ChIP-seq peaks are available from the ENCODE Project (www.encodeproject.org). K562 ATAC-seq peaks are available from the Gene Expression Omnibus (GEO) under accession number GSE161002; K562 GFI1B ChIP-seq peaks are available from GEO under accession number GSE117944.

Plasmid cloning for lentiviral CRISPRi, cytosine base editor, and modified gRNA scaffold vectors

To generate the KRAB-dCas9 (lentiCRISPRi (v1)-Blast) and KRAB-dCas9-MeCP2 [lentiCRISPRi(v2)-Blast] plasmids, KRAB and dCas9 were polymerase chain reaction (PCR) amplified from pCC_09 (Addgene 139094) (*92*), and the MeCP2 effector domain was synthesized as a gBlock (IDT). KRAB and MeCP2 were linked to dCas9 with flexible glycine-serine linkers and cloned into lentiCas9-Blast (Addgene 52962) (*23*). To generate the FNLS-BE3-SpRY (lentiBE3-SpRY-Blast) plasmid, we used Gibson cloning to replace the puromycin resistance gene in pLenti-FNLS-P2A-Puro (Addgene 110841) with blasticidin resistance from lentiCRISPRi(v2)-Blast. We then used Gibson cloning to replaced SpCas9(D10A) with the SpRY nickase from pCAG-CBE4max-SpRY-P2A-EGFP (Addgene 139999) (*51*). To generate the gRNA vector (lentiGuideFE-Puro), we digested pCC_09 with NheI and KpnI to isolate the U6 promoter and Cas9 guide RNA scaffold with the F+E scaffold modification (*93*). After gel extraction (Qiagen 28706), we ligated this piece into NheI- and KpnI-digested pLentiRNAGuide_001 (Addgene 138150) vector using T4 ligase (NEB M0202M) (*94*). Primer sequences for Gibson cloning reactions are available in table S1F.

Cell culture and monoclonal cell line generation

Human embryonic kidney (HEK) 293FT cells were acquired from Thermo Fisher Scientific (R70007). HEK293FT (human) cells were maintained at 37°C with 5% CO_2 in D10 medium: Dulbecco's modified Eagle's medium (DMEM) with high glucose and stabilized L-glutamine (Caisson DML23) supplemented with 10% fetal bovine serum (FBS) (Thermo Fisher 16000044). K562 cells were acquired from ATCC (CCL-243) and were maintained at 37°C with 5% CO_2 in R10 medium: RPMI with stabilized L-glutamine (Thermo Fisher 11875119) supplemented with 10% FBS (Thermo Fisher 16000044). Cells were regularly passaged and tested for the presence of mycoplasma contamination with the MycoAlert Plus Mycoplasma Detection Kit (Lonza).

Lentivirus was produced by polyethylenimine linear MW 25000 (Polysciences 23966) transfection of HEK293FT cells with the transfer plasmid containing a Cas9 effector, or gRNA library, packaging plasmid psPAX2 (Addgene 12260) and envelope plasmid pMD2.G (Addgene 12259). At 72 hours after transfection, cell medium containing lentiviral particles was harvested and filtered through 0.45-mm filter Steriflip-HV (Millipore SE1M003M00). K562 cells were transduced with lentiCRISPRi(v1)-Blast, lentiCRISPRi(v2)-Blast, or lentiBE3-SpRY–Blast at a low MOI (<1). Transduced K562 cells were selected with 10 μg/μl blasticidin (Thermo A1113903) for 10 days to enrich for expression of the Cas9 effector proteins. To isolate individual clones, K562 polyclonal lines were serially diluted to 50 cells per 10 ml of medium. We then plated 100 μl of this cell-medium mixture in 96-well round bottom plates (~0.5 cells/well).

Digital PCR for CRISPRi gene repression

We compared the single-repressor CRISPRi (KRAB-dCas9) and dual-repressor CRISPRi (KRAB-dCas9-MeCP2) systems by targeting the transcription start sites and known enhancers of three genes (*MRPS23*, *SLC25A37*, and *FSCN1*) with two gRNAs per targeted region. We synthesized gRNAs as top- and bottom-strand oligos (IDT) and cloned them into BsmBI-digested lentiGuideFE-Puro. We transduced the cells in biological triplicate with gRNA lentiviruses at a low MOI and after 24 hours selected for cells with gRNAs using puromycin (1 μg/μl, Thermo Fisher A1113803). We harvested the cells 10 days after transduction and extracted RNA using TRIzol (ThermoFisher 15596026). We quantified RNA concentration by spectrophotometry (NanoDrop). To measure gene expression, we performed digital PCR (Formulatrix Consellation) with Cy5/Iowa Black RQ target gene probes (IDT), FAM/ZEN/Iowa Black FQ for the actin normalizer (IDT), and Luna Universal One-Step RT qPCR Master Mix kit (NEB E3005L) and Tween-20 (Sigma-Aldrich P1379). We first normalized the target gene expression by actin expression per sample and then normalized this ratio to the ratio from cells transduced with nontargeting control gRNAs.

KRAB-dCas9-MeCP2 CRISPRi pooled screen for essential gene gRNA depletion

We performed CRISPRi pooled screens to quantify the KRAB-dCas9-MeCP2 inhibitory effect window in HCT116 and MCF7 cell lines. Both lines were acquired from ATCC (CCL-247 and and HTB-22, respectively) maintained in the appropriate medium (McCoy's 5A medium and DMEM, respectively) supplemented with 10% serum and 1% penicillin–streptomycin. These cell lines were cultured at 37°C, 5% CO_2, and ambient oxygen levels. Monoclonal HCT116 KRAB-dCas9-MeCP2 and MCF7 KRAB-dCas9-MeCP2 cell lines were generated as previously described for K562 cells. Expression was confirmed using Western blot.

For screening, HEK293 cells were plated in DMEM + 10% FBS (D10) in a 15-cm dish so that the following day cells were 90% confluent. Half of the medium was removed from the flask, and cells in each flask were transfected with 13.8 μg of a cCRE/TSS–targeting library specific to HCT116 and MCF7, 6.6 μg of pMD2.G (envelope plasmid), and 9.6 μg of psPAX2 (packaging plasmid) using 1.2 ml of Opti-MEM and 112.5 μl of polyethylenimine linear 25K (Polysciences 23966). The following morning, the medium was removed and fresh D10 + 1% bovine serum albumin (BSA) was added. Then, 48 hours later, we collected the viral supernatant and put it immediately on ice. We concentrated the supernatant by centrifugation at 100,000*g* (Thermo Sorvall LYNX) for 2 hours at 4°C. The resulting pellet was resuspended in cold DMEM and stored at –80°C until use.

We determined the appropriate titer of virus before the experimental transduction. We transduced 3M cells with a standard spinfection protocol with different dilutions of virus in a 12-well plate and in a no-virus control well. After adding virus, we spun the cells at 2000 rpm for 1 hour at 37°C (Beckman Coulter Allegra X-14R) and incubated overnight. The next day, we plated half of the cells in each well into two new wells of a six-well plate. In one set of wells, we added the appropriate puromycin concentration (1.5 μg/ml for HCT116 and 3 μg/ml for MCF7). After all the cells in the no-virus well had died, cells in the corresponding wells (with puromycin) were counted to determine the viral volume that results in 20 to 40% cell survival, corresponding to a MOI of 0.2 to 0.5.

We cultured each cell line in the appropriate medium and transduced 2×10^8 of them with the CRISPR lentiviral library using spinfection with the viral volume determined from the previous spinfection. As before, after adding virus, we spun cells at 2000 rpm for 1 hour at 37°C and incubated them overnight. The following day, cells were plated at 30% confluence and selected in the appropriate puromycin concentration for 3 days. After selection, we passaged cells in 15-cm dishes for 21 days and split at ~80% confluence. We isolated genomic DNA from cells using a modified salting-out precipitation. The gRNA readout was performed using two rounds of PCR. For PCR1, we used 10 μg of gDNA in each 100-μl reaction. We pooled the PCR1 products and used the mixture for a second PCR, adding on Illumina sequencing adaptors and barcodes. We performed PCR1 reactions using TaqB polymerase (Enzymatics P7250L) and PCR2 reactions with Q5 (NEB M0491). We pooled and purified PCR2 reactions with Illumina Purification Beads. We quantified the concentration of the gel-extracted PCR products using Qubit dsDNA HS Assay Kit (Thermo Fisher Q32851), and then diluted and sequenced it on NextSeq 500 high-output (Illumina). We demultiplexed the samples using bcl2fastq v2.20.0.422 (Illumina), trimmed off adapters, and aligned to our guides with bowtie v.1.1.2 (*95*). We library normalized the resulting reads (each read divided by the total number of reads). We then used the robust rank aggregation algorithm (*96*) and

estimataed $\log_2$ fold changes as $\log_2$(day 21/day 1). We targeted ± 5 kb of the TSS essential genes (DepMap Chronos scores < –1) (*97–100*). In total, we screened 1992 gRNAs targeting 263 essential human genes. As negative controls, we embedded 1000 nontargeting gRNAs into this library.

Flow sorting for near PAM-less base editing

We verified cytosine base editing by designing 12 gRNAs targeting *CD46* splice sites using SpliceR v1.2.0 (*101*). SpliceR designed gRNAs that were predicted to disrupt *CD46* splice sites through C>T nucleotide changes. These included gRNAs that would recognize a diverse set of noncanonical PAMs, such as NGN, NAN, NCN, and NTN (table S3H). We also used four nontargeting gRNAs from the GeCKOv2 library (*23*) as negative controls. We synthesized gRNAs as top- and bottom-strand oligos (IDT) and cloned them into BsmBI-digested lentiGuideFE-Puro. We transduced the cells with gRNA lentiviruses at a low MOI in an arrayed fashion and after 24 hours selected for cells with gRNAs using puromycin (1 μg/μl, Thermo Fisher A1113803). After 6 days of selection, we proceeded to flow cytometry to measure CD46 protein. For flow cytometry, 1×10^6 cells per condition were harvested and washed with phosphate-buffered saline (PBS) after selection. The cells were stained for 5 min at room temperature with LIVE/DEAD Fixable Violet Dead Stain Kit (Thermo Fisher L34864). Subsequently, the cells were stained with antibodies for 20 min on ice with 1 μl of CD46-APC (clone TRA-2-10) (BioLegend 352405). Cells were washed with PBS to remove unbound antibodies before sorting. Cell acquisition and sorting was performed using a Sony SH800S cell sorter. Sequential gating was performed as follows: exclusion of debris on the basis of forward and side scatter cell parameters followed by dead cell exclusion. The sorting gates were set such that 90% of live K562 cells would be considered CD46 positive.

CRISPR inhibition and base-editing library design and cloning

Two individual CRISPR inhibition libraries were designed and cloned, called STING-seq v1 and STING-seq v2, and one base-editing library was designed and cloned, called beeSTING-seq. For STING-seq v1, we designed 20-nt gRNAs to target within 200 base pairs (bp) of the 88 selected plausibly causal noncoding variants from UKBB GWASs of blood traits. We used FlashFry v1.10.0 (*22*) to retain gRNAs with the lowest predicted off-target activity, as estimated by the Hsu-Scott score (*21*). Each variant was targeted by two different gRNAs. In addition, we included in our library 12 nontargeting gRNAs from the GeCKOv2 library (*23*) as negative controls and 12 gRNAs targeting the TSSs of six nonessential genes as positive controls. The six nonessential genes (*CD46*, *CD52*, *HSPA8*, *NMU*, *PPIA*, and *RPL22*) were identified by a CRISPR knock-out screen in K562 cells using the PICKLES database (*102*). We additionally included 10 gRNAs targeting the CD55 TSS for our fluorescence-activated cell sorting (FACS)–based MOI estimator, bringing the total number of gRNAs to 210. For STING-seq v2, we designed 20-nt gRNAs to target within 200 bp of the 507 selected plausibly causal variants from the BCX multiancestry and ancestry-specific blood trait GWASs. We again retained gRNAs with the lowest predicted off-target activity, and each variant was targeted by three different gRNAs. In addition, we included 30 nontargeting gRNAs from the GeCKOv2 library and 32 groups of three TSS-targeting gRNAs for positive controls. We additionally included 45 CD55 TSS-targeting gRNAs for FACS-based MOI estimation. For beeSTING-seq, we designed three sets of gRNAs for each of 46 C>T select GWAS variants mapping to CREs with *cis*-target genes. We followed recommended gRNA design instructions, and positioned the target nucleotide within a 5-nt window (*103*). We also included 28 nontargeting gRNAs from the GeCKOv2 library.

To clone the STING-seq v1 gRNA library, top- and bottom-strand oligos (IDT) were resuspended in water at 100 μM and then mixed at 1:1 ratio for each gRNA. Then, 1 μl of the oligo mix was added to a master mix containing 1× T4 ligase buffer (NEB M0202M) and 0.5 μl of T4 PNK (NEB M0201L) and water to a final concentration of 10 μl per reaction. For oligo annealing, we incubated the oligo mix at 37°C for 30 min, then 95°C for 5 min with a temperature change of 1°C every 5 s until reaching 4°C. To create the oligo pool, we pooled together 3 μl of each annealed oligo. The oligo pool was diluted 1:10 with water and then cloned in the lentiGuideFE-Puro, which was linearized with BsmBI (Thermo Fisher ER0451) and dephosphorylated. The ligation was performed in 11 reactions, with each reaction consisting of 5 μl of Rapid Ligation Buffer (Enzymatics B101), 0.5 μl of T7 ligase (Enzymatics L602L), digested plasmid at 25 ng per reaction, 1 μl of diluted oligo mix and double-distilled water to final volume of 10 μl. The ligation was performed at room temperature for 15 min.

Next, 100 μl of the combined ligation reactions were mixed with 100 μl of isopropanol, 1 μl of GlycoBlue (Thermo Fisher AM9515), and 2 μl of 5 M NaCl (50 mM final concentration), incubated for 15 min at room temperature, and spun at 12,000*g* for 15 min. The pellet was washed twice with prechilled 70% ethanol, air dried for 15 min or until dried completely, and resuspended in 5 μl of 1× TE buffer (Sigma). Next, 2 μl of library ligation was added to 50 μl of Endura cells (Lucigen) and then electroporated, recovered, and plated. The following day, bacterial colonies were scraped, plasmids were isolated using a maxi prep (Qiagen 12965), and library representation was determined by MiSeq (Illumina).

The STING-seq v2 and beeSTING-seq pooled gRNA libraries were synthesized as single-stranded oligonucleotide pools (Twist Biosciences) and diluted to 0.5 ng/μl in molecular-grade water. Then, 2 μl of the diluted pooled oligos were added to a master mix containing forward and reverse primer mixes (10 μM) and NEBNext High-Fidelity 2X PCR Master Mix (M0541S). We then PCR purified the product and Gibson cloned it in pLentiGuideFE-Puro, which was linearized as described above. We used 500 ng of the digested vector, maintained a 1:10 molar ratio of library, and incubated at 50°C for 1 hour. We concentrated DNA using isopropanol precipitation, washed and resuspended the DNA, and then transformed 1 μl of library in 25 μl of Endura cells (Endura 60242-2) according to protocol specifications. We then plated the transformed cells on Luria broth–ampicillin plates to get at least 100 to 500 colonies per gRNA.

The quality of all pooled libraries was verified by sequencing with a MiSeq (Illumina) to estimate the 90:10 quantile ratio. To generate and concentrate all pooled libraries, lentivirus was generated as described above. Briefly, we seeded 10 × 225 cm^2 flasks with HEK293FT cells and, at 70% confluency, we cotransfected the pooled gRNA library, psPAX2, and pMD2.G. Lentivirus was collected 72 hours after transfection and filtered using a 0.45-μm filter. The supernatant was then ultracentrifuged for 2 hours at 100,000*g* (Sorvall Lynx 6000), and the pellet was resuspended overnight at 4°C in PBS with 1% BSA.

MOI estimation using flow cytometry

When transducing cells at a high MOI, it is not possible to estimate the MOI by traditional methods (e.g., survival after drug selection) or without the time and cost of single-cell sequencing. By including multiple gRNAs that target the *CD55* TSS (10 gRNAs for STING-seq v1, 45 gRNAs for STING-seq v2), we were able to estimate the number of gRNAs per cell (MOI) using flow cytometry for CD55 cell surface protein knockdown over a range of viral transduction volumes. We performed two transductions for STING-seq v1 with concentrated lentivirus (4 and 6 μl) and, after 48 hours, we selected with puromycin for 10 days. We performed five transductions for STING-seq v2 with concentrated lentivirus (1, 5, 10, 20, and 30 μl) and, after 48 hours, we selected with puromycin for 10 days. We included three positive control transductions with different *CD55* TSS-targeting gRNAs and three negative control transductions with three different nontargeting gRNAs for both experiments. For beeSTING-seq, we performed five transductions with concentrated lentivirus (1, 5, 10, 25, and

50 μl), and, after 48 hours, we selected with puromycin for 10 days. We used the most viable cell culture for beeSTING-seq for sequencing (10 μl) with MACS dead cell removal kit (Miltenyi Biotec 130-090-101) because we observed high cell death at higher lentivirus concentrations.

For flow cytometry, 1×10^6 cells per condition were harvested and washed with PBS after selection. The cells were stained for 5 min at room temperature with LIVE/DEAD Fixable Violet Dead Stain Kit (Thermo Fisher L34864). Subsequently, the cells were stained with antibodies for 20 min on ice with 1 μl of CD55-FITC (clone JS11) (BioLegend 311306). Cells were washed with PBS to remove unbound antibodies before sorting. Cell acquisition and sorting was performed using a Sony SH800S cell sorter. Sequential gating was performed as follows: exclusion of debris on the basis of forward and side scatter cell parameters followed by dead cell exclusion. The sorting gates were set such that 90% of live K562 cells would be considered CD55 positive. From this estimation, we can estimate MOI using $X = 1 - N^Y$, where X is the proportion of cells with CD55 targeting gRNAs, N is the inverse of the number of CD55 targeting gRNAs divided by the total library size, leaving X as the predicted MOI. For the STING-seq v1 library, $N = 1 - (10/210)$, and for the STING-seq v2 library, $N = 1 - (45/1695)$. We estimated that 6 μl of STING-seq v1 viral volume yielded an MOI of ~13.5 and 30 μl of STING-seq v2 viral volume yielded an MOI of ~30, and elected to use these conditions for our STING-seq assay (fig. S2).

Expanded CRISPR-compatible Cellular Indexing of Transcriptomes and Epitopes

For the Expanded CRISPR-compatible Cellular Indexing of Transcriptomes and Epitopes sequencing (ECCITE-seq) and the STING-seq v1 experiment, we ran one lane of a 10× Genomics 5′ kit (Chromium Single Cell Immune Profiling Solution v1.0, 1000014, 1000020, and 1000151) with superloading and recovered 15,285 total cells (including multiple cells per droplet counts or "multiplets"). Cell hashing was performed as described in a previously published protocol using four hashtag-derived oligonucleotides (HTOs) using hyperconjugation (*24*). Gene expression (cDNA), hashtags (HTOs), and gRNA (guide-derived oligos, GDOs) libraries were constructed by following 10x Genomics and ECCITE-seq protocols. We sequenced the cDNA, HTO, and GDO libraries with two NextSeq 500 high-output runs (Illumina). For the ECCITE-seq and the STING-seq v2 experiment, we ran four lanes of a 10x Genomics 5′ v2 kit (Chromium Next GEM Single Cell 5′ Kit v2 1000265) with superloading. We recovered 82,339 total cells (including multiplets). Cell hashing was performed using eight HTOs followed by staining with a 188 antibody-tagged oligonucleotides (ADTs) panel (BioLegend) (table S3B). The cDNA, HTO, ADT, and GDO libraries were constructed by following 10x Genomics and ECCITE-seq protocols. We sequenced the cDNA, HTO, ADT, and GDO libraries with one NovaSeq 6000 S1 run and two NovaSeq 6000 S2 runs (Illumina). For the ECCITE-seq and beeSTING-seq experiment, we ran three lanes of a 10x Genomics 5′ v2 kit with superloading and recovered 39,049 total cells, including multiplets. Cell hashing was performed using nine HTOs. The cDNA, HTO, and GDO libraries were constructed by following 10x Genomics and ECCITE-seq protocols. We sequenced the cDNA, HTO, and GDO libraries with one NextSeq 500 mid-output run, one NovaSeq 6000 SP run, and one NovaSeq 6000 S1 run (Illumina).

Single-cell data processing

UMI count matrices were generated for all single-cell sequencing libraries with 10x Cell Ranger v.6.0.0 (*104*). We generated outputs using the Gene Expression Output, Antibody Capture Output, and CRISPR Guide Capture Output functions. We then analyzed the UMI count matrices in R v.4.0.2 with Seurat v.4.0.0 (*105*) and tested for differential gene expression and protein levels within the SCEPTRE framework (*26*). The distributions of cDNA, GDO, HTO, and ADT UMIs were inspected manually for each lane of 10× sequenced. Custom thresholds were set to remove outliers for total cDNA count, unique genes detected, mitochondrial percentage, total gRNA count, unique gRNAs detected, total HTO count, unique HTOs detected, total ADT count, and unique ADTs detected. Lanes were merged for STING-seq v2 and beeSTING-seq only after quality control was completed. For STING-seq v1, we processed cDNA UMI count matrices and retained cells between the 15th to 99th percentiles for unique gene count, between the 20th and 99th percentiles for total cDNA UMI count, and between the 5th and 90th percentile for mitochondrial percentage. Next, we center-log-ratio (CLR) transformed the HTO UMI counts and demultiplexed cells by their transformed HTO counts to identify singlets. We used the HTODemux function implemented in Seurat v.4.0.0 to maximize the number of singlets detected. We used then processed the GDO UMI count matrix, keeping cells between the 1st and 99th percentiles for total GDO count and used 10x Cell Ranger predicted GDO thresholds per cell, but required at least three UMIs per GDO to assign a GDO to a given cell. This resulted in a high-confidence set of 7667 single cells for the STING-seq v1 experiment. For STING-seq v2, we uniformly processed all four cDNA UMI count matrices and retained cells between the 1st and 99th percentile for unique gene count, between the 10th and 99th percentile for total cDNA UMI count, and between the 1st and 90th percentile for mitochondrial percentage. Next, we CLR transformed the HTO UMI counts and maximized singlet count using the HTODemux function. We then processed the GDO UMI count matrices, keeping cells between the 1st and 99th percentiles for total GDO count and again used the 10x Cell Ranger predicted GDO thresholds per cell, but required at least three UMIs per GDO. This resulted in a high-confidence set of 38,916 cells for differential expression testing. We further applied quality control filters for ADTs, retaining cells with between the 1st and 99th percentiles for total ADT count. This resulted in 38,133 cells for differential protein testing. For beeSTING-seq, we uniformly processed all three cDNA UMI count matrices and retained cells between the 10th and 90th percentiles for unique gene count, between the 10th and 90th percentiles for total cDNA count, and between the 10th and 90th percentiles for mitochondrial percentage. We then CLR transformed the HTO counts and used the HTODemux function to maximize singlets and retained cells between the 1st and 99th percentiles for total GDO counts. 10x Cell Ranger set most UMI thresholds to 1, so we generated a series of GDO UMI count matrices with thresholds from 1 to 5 to iteratively test optimal GDO thresholds for each gRNA. This resulted in a series of UMI count matrices for each GDO threshold. We had sets of 12,068 cells (GDO threshold = 1), 11,235 cells (GDO threshold = 2), 9739 (GDO threshold = 3), 7869 (GDO threshold = 4), and 5896 (GDO threshold = 5) for differential expression testing.

Differential gene expression– and protein-level testing with SCEPTRE

We used the processed UMI count matrices for gene expression or protein levels and gRNA expression, along with accompanying single-cell metadata to use as covariates in model fitting (table S3B). For STING-seq analyses, we defined for each cCRE targeted by two to three gRNAs a list of genes within 500 kb to be tested for differential expression in *cis*. For each gene per set of gRNAs, we extracted that gene's UMI counts and labeled the cells with the given gRNAs. We then tested for differential outcomes within the SCEPTRE framework (*26*), adjusting for the following single-cell covariates for expression tests: total gene expression UMIs, unique genes, total gRNA expression UMIs, unique gRNAs, percentage of mitochondrial genes, and 10x lane (for STING-seq v2 and beeSTING-seq). For protein tests, we adjusted for total ADT count, total HTO count, total gRNA expression UMIs, unique gRNAs, and ADTs for four mouse-specific antibody controls to represent nonspecific binding. We developed SCEPTRE as a statistical framework to analyze high-MOI CRISPR screens in single cells with

state-of-the-art calibration. First, SCEPTRE fits a negative binomial distribution to measure the effect of a single gRNA on a given gene using the Z score. Then, the distribution of gRNAs to cells is randomly sampled to build a gRNA-specific null distribution, recomputing a negative binomial Z score. A skew-t distribution is fit to compare the test Z score and the null distribution, and a two-sided P value is derived, allowing for significance tests of increased or decreased gene expression or protein levels (*26*). To test for differential expression in *trans*, we defined for each set of gRNAs a list of all genes detected in at least 5% of cells and repeated the test above. Nontargeting gRNAs were tested against all genes used in the *cis* and *trans* settings discussed previously and randomly sampled to match the number of *cis* and *trans* tests displayed on QQ-plots. For each set of gRNAs with a *cis*-effect target gene we then performed marginal gRNA-gene pair testing, observing that the number of cells bearing gRNAs is the main driver behind statistical power (fig. S6). To determine significance for multiple hypotheses (genes) tested in *cis*, SCEPTRE P values were adjusted with the Benjamini-Hochberg procedure. For beeSTING-seq differential expression tests, we tested each gRNA against its known target gene from STING-seq analyses. We used for each gRNA the lowest GDO UMI threshold that resulted in at least 100 cells per gRNA and repeated this strategy for all nontargeting gRNAs against the same set of known *cis*-effect target genes. We then repeated differential expression testing, grouping together GWAS-CRE targeting gRNAs if they shared concordant effects and UMI thresholds and evaluating their combined effects on target gene expression.

To report significant results for STING-seq analyses, we identified *cis*-target genes if they were significant at a 5% FDR (Benjamini-Hochberg–adjusted SCEPTRE $P < 0.05$). We defined *trans*-target genes of each GWAS-CRE as those significant at a stricter 1% FDR. For beeSTING-seq analyses, we identified *cis*-target genes if they were significant at a 5% FDR. We examined all STING-seq genes significant at a 10% FDR and beeSTING-seq genes with SCEPTRE $P < 0.05$ to compare the *trans*-regulatory network effects from perturbing *HHEX* and *IKZF1* GWAS-CREs with direct variant insertion.

Fine-mapped eQTL credible set integration

We examined 31 fine-mapped eQTL studies from the eQTL Catalogue (*31*) specific to blood traits. Specifically, we used eQTLs identified from human macrophages (*106*, *107*), monocytes (*37*, *108*, *109*), neutrophils (*37*), lymphoblastoid cell lines (*110*–*113*), whole blood (*110*, *114*, *115*), induced pluripotent stem cells (*116*–*118*), T cells (*37*, *109*, *112*), B cells (*109*), and natural killer cells (*109*). We then retained eQTL variants that were at least 1% plausibly causal and investigated whether our fine-mapped GWAS variants were in these data. eQTL summary statistics are available from the eQTL Catalog (www.ebi.ac.uk/eqtl).

K562 HiChIP for H3K27ac-interacting promoters

AQuA-HiChIP cell libraries were prepared as described previously (*119*). Briefly, NIH3T3 cells (mouse) and K562 cells were grown in the appropriate medium. Cells were fixed in 1% formaldehyde for 10 min and quenched to a final concentration of 125 nM glycine. Two million fixed mouse cells were mixed with 10 million fixed K562 cells. The cells were lysed in 0.5% SDS, quenched with 10% Triton X-100, and digested with MboI (NEB R0147M). The DNA overhangs were blunted, biotinylated (Thermo Fisher 19524016), and ligated. Nuclei were spun down, resuspended in nuclear lysis buffer, and sonicated using a Covaris LE220 with the following conditions: fill level 10, PIP 450, duty factor 30, and CPB 200. The sheared DNA was incubated with Dynabeads Protein A (Thermo Fisher 10001D) for 2 hours at 4°C. The tubes were placed on a magnet and the supernatant was kept. Immunoprecipitation was performed with a cross-species reactive H3K27ac antibody (Active Motif 39133). The samples were incubated with the antibody overnight at 4°C. The samples were then washed, eluted, and treated with Proteinase K. The samples were purified using Zymo DNA Clean & Concentrator. Biotin capture was performed with Dynabeads M-280 Streptavidin (Thermo Fisher 11205D), followed by library preparation. The amplified libraries were purified with Illumina Sample Purification Beads. The libraries were sequenced using paired-end reads with a NovaSeq 6000 S2 (Illumina) to generate 100 to 200 million read pairs per sample.

HiChIP paired end reads were mapped to the hg19 genome using HiC-Pro v.2.10.1 (*120*). Default settings were using to remove duplicate reads, identify valid interactions, and generate contact maps. Statistically significant contacts were identified using FitHiChIP v.9.1 (*121*) at a 5% FDR. H3K27ac ChIP-seq data (*65*) were used as a reference set of peaks in the FitHiChIP pipeline.

Trans-regulated network gene set enrichments

We used chromatin immunoprecipitation sequencing (ChIP-seq) datasets in K562 cells to identify *GFI1B* (*64*), *NFE2*, *IKZF1*, and *RUNX1* (*65*) TF-binding sites. There we no publicly available *HHEX* K562 ChIP-seq datasets. We assigned the closest protein-coding gene to each ChIP-seq peak with bedtools v2.25.0 (*90*). For predicted miRNA targets we used the TargetScan database (*66*, *67*). To test for enrichment of ChIP-seq peak or TargetScan genes in *trans*-regulatory gene sets, we fit logistic regression models adjusting for K562 expression (gene expression counts from scRNA-seq data) and computed odds ratios with 95% confidence intervals. To construct GWAS-identified sets of genes, we used all fine-mapped SNPs from the 29 UKBB GWASs and 15 BCX GWASs previously described (categorized by cell type) with a high Bayes factor for being plausibly causal ($\log_{10}$ BF ≥ 2) and that were at least 1% plausibly causal. GWAS gene enrichment was performed in a similar fashion as for ChIP-seq peaks.

Gene coexpression analyses and bone marrow single-cell gene expression

To compute coexpression matrices for each *trans*-regulatory network, we used cDNA UMI count matrices with missing genes per cell imputed with the MAGIC algorithm (*122*). As a measure of coexpression, the biweight mid-correlation, a weighted correlation analysis, was calculated for each pair of genes (*123*). Genes were then clustered on the basis of their coexpression patterns by hierarchical clustering. TF-binding site, direct miRNA target, and GWAS gene enrichment was performed as described above. We used Human Cell Atlas single-cell RNA-sequencing from 35 bone marrow donors (*69*) and identified 27 cell types as described previously (*70*). Single-cell data were processed with Seurat v.4.0.0 to generate uniform manifold approximation and projection (UMAP) plots and heatmaps. To visualize entire *trans*-regulatory network clusters on a UMAP plot, we plotted the mean expression of all cluster genes within each cell.

STING-seq power estimations

We down-sampled 136 *cis* effects of gRNAs targeting CREs on their target genes across two key conditions for experimental design: sequencing read depth per cell and the number of cells per gRNA. We sequenced all STING-seq libraries to a depth of ~55,000 to 65,000 reads per cell and thus repeated the entire STING-seq quality control and differential expression testing pipeline with 5000, 15,000, 25,000, 35,000, 45,000 and 55,000. Sequencing reads were down-sampled to generate cDNA UMI count matrices with DropletUtils v.1.18.0 (*124*, *125*) and repeated 10 times with different seed numbers. For each set of 10 randomly down-sampled UMI count matrices at each read depth, we repeated differential expression testing with SCEPTRE. We required at least 500 cells bearing each set of gRNAs, then at each set of 10 randomly down-sampled UMI count matrices at each read depth, we randomly down-sampled the number of cells bearing each set of gRNA from at least 500 cells to 50 and repeated this process 10 times at each stage. We averaged the SCEPTRE skew fit t test P values within replicates at each to compute precise measurements

for each stage in the down-sampling procedures. We then divided all genes by their expression level and *cis* effects by their $\log_2$-fold changes into tertiles to examine at what number of cells and read depth could nominal significance (skew fit t test $P < 0.0001$) be attained.

REFERENCES AND NOTES

1. J. B. Wright, N. E. Sanjana, CRISPR screens to discover functional noncoding elements. *Trends Genet.* **32**, 526–529 (2016). doi: 10.1016/j.tig.2016.06.004; pmid: 27423542
2. T. Lappalainen, D. G. MacArthur, From variant to function in human disease genetics. *Science* **373**, 1464–1468 (2021). doi: 10.1126/science.abi8207; pmid: 34554789
3. J. C. Ulirsch *et al.*, Interrogation of human hematopoiesis at single-cell and single-variant resolution. *Nat. Genet.* **51**, 683–693 (2019). doi: 10.1038/s41588-019-0362-6; pmid: 30858613
4. J. A. Morris *et al.*, An atlas of genetic influences on osteoporosis in humans and mice. *Nat. Genet.* **51**, 258–266 (2019). doi: 10.1038/s41588-018-0302-x; pmid: 30598549
5. J. Nasser *et al.*, Genome-wide enhancer maps link risk variants to disease genes. *Nature* **593**, 238–243 (2021). doi: 10.1038/s41586-021-03446-x; pmid: 33828297
6. Q. Sun *et al.*, From GWAS variant to function: A study of ~148,000 variants for blood cell traits. *HGG Adv.* **3**, 100063 (2021). doi: 10.1016/j.xhgg.2021.100063; pmid: 35047852
7. M. Gasperini *et al.*, A genome-wide framework for mapping gene regulation via cellular genetic screens. *Cell* **176**, 377–390.e19 (2019). doi: 10.1016/j.cell.2018.11.029; pmid: 30612741
8. S. Xie, D. Armendariz, P. Zhou, J. Duan, G. C. Hon, Global analysis of enhancer targets reveals convergent enhancer-driven regulatory modules. *Cell Rep.* **29**, 2570–2578.e5 (2019). doi: 10.1016/j.celrep.2019.10.073; pmid: 31775028
9. F. Wünnemann *et al.*, Multimodal CRISPR perturbations of GWAS loci associated with coronary artery disease in vascular endothelial cells. *PLOS Genet.* **19**, e1010680 (2023). doi: 10.1371/journal.pgen.1010680; pmid: 36928188
10. W. J. Astle *et al.*, The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell* **167**, 1415–1429.e19 (2016). doi: 10.1016/j.cell.2016.10.042; pmid: 27863252
11. M.-H. Chen *et al.*, Trans-ethnic and ancestry-specific blood-cell genetics in 746,667 individuals from 5 global populations. *Cell* **182**, 1198–1213.e14 (2020). doi: 10.1016/j.cell.2020.06.045; pmid: 32888493
12. D. Vuckovic *et al.*, The polygenic and monogenic basis of blood traits and diseases. *Cell* **182**, 1214–1231.e11 (2020). doi: 10.1016/j.cell.2020.08.008; pmid: 32888494
13. J. Yang, S. H. Lee, M. E. Goddard, P. M. Visscher, GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011). doi: 10.1016/j.ajhg.2010.11.011; pmid: 21167468
14. J. Yang *et al.*, Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375, S1–S3 (2012). doi: 10.1038/ng.2213; pmid: 22426310
15. V. G. Sankaran *et al.*, Cyclin D3 coordinates the cell cycle during differentiation to regulate erythrocyte size and number. *Genes Dev.* **26**, 2075–2087 (2012). doi: 10.1101/gad.197020.112; pmid: 22929040
16. J. C. Ulirsch *et al.*, Altered chromatin occupancy of master regulators underlies evolutionary divergence in the transcriptional landscape of erythroid differentiation. *PLOS Genet.* **10**, e1004890 (2014). doi: 10.1371/journal.pgen.1004890; pmid: 25521328
17. J. C. Ulirsch *et al.*, Systematic functional dissection of common genetic variation affecting red blood cell traits. *Cell* **165**, 1530–1545 (2016). doi: 10.1016/j.cell.2016.04.048; pmid: 27259154
18. J. Wen *et al.*, Super interactive promoters provide insight into cell type-specific regulatory networks in blood lineage cell types. *PLOS Genet.* **18**, e1009984 (2022). doi: 10.1371/journal.pgen.1009984; pmid: 35100265
19. N. C. Yeo *et al.*, An enhanced CRISPR repressor for targeted mammalian gene regulation. *Nat. Methods* **15**, 611–616 (2018). doi: 10.1038/s41592-018-0048-5; pmid: 30013045
20. L. A. Gilbert *et al.*, Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* **159**, 647–661 (2014). doi: 10.1016/j.cell.2014.09.029; pmid: 25307932
21. P. D. Hsu *et al.*, DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013). doi: 10.1038/nbt.2647; pmid: 23873081
22. A. McKenna, J. Shendure, FlashFry: A fast and flexible tool for large-scale CRISPR target design. *BMC Biol.* **16**, 74 (2018). doi: 10.1186/s12915-018-0545-0; pmid: 29976198
23. N. E. Sanjana, O. Shalem, F. Zhang, Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014). doi: 10.1038/nmeth.3047; pmid: 25075903
24. M. Stoeckius *et al.*, Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017). doi: 10.1038/nmeth.4380; pmid: 28759029
25. E. P. Mimitou *et al.*, Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. *Nat. Methods* **16**, 409–412 (2019). doi: 10.1038/s41592-019-0392-0; pmid: 31011186
26. T. Barry, X. Wang, J. A. Morris, K. Roeder, E. Katsevich, SCEPTRE improves calibration and sensitivity in single-cell CRISPR screen analysis. *Genome Biol.* **22**, 344 (2021). doi: 10.1186/s13059-021-02545-2; pmid: 34930414
27. R. Andersson *et al.*, An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014). doi: 10.1038/nature12787; pmid: 24670763
28. E. B. Fauman, C. Hyde, An optimal variant to gene distance window derived from an empirical definition of cis and trans protein QTLs. *BMC Bioinformatics* **23**, 169 (2022). doi: 10.1186/s12859-022-04706-x; pmid: 35527238
29. D. Yao *et al.*, Multi-center integrated analysis of non-coding CRISPR screens. bioRxiv 520137 [Preprint] (2022); https://doi.org/10.1101/2022.12.21.520137.
30. GTEx Consortium; Laboratory, Data Analysis & Coordinating Center (LDACC)—Analysis Working Group; Statistical Methods groups—Analysis Working Group, Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017). doi: 10.1038/nature24277; pmid: 29022597
31. N. Kerimov *et al.*, A compendium of uniformly processed human gene expression and splicing quantitative trait loci. *Nat. Genet.* **53**, 1290–1299 (2021). doi: 10.1038/s41588-021-00924-w; pmid: 34493866
32. B. Rowland *et al.*, Transcriptome-wide association study in UK Biobank Europeans identifies associations with blood cell traits. *Hum. Mol. Genet.* **31**, 2333–2347 (2022). doi: 10.1093/hmg/ddac011; pmid: 35138379
33. B. Li *et al.*, Tissue specificity-aware TWAS (TSA-TWAS) framework identifies novel associations with metabolic, immunologic, and virologic traits in HIV-positive adults. *PLOS Genet.* **17**, e1009464 (2021). doi: 10.1371/journal.pgen.1009464; pmid: 33901188
34. M. T. Maurano *et al.*, Large-scale identification of sequence variants influencing human transcription factor occupancy in vivo. *Nat. Genet.* **47**, 1393–1401 (2015). doi: 10.1038/ng.3432; pmid: 26502339
35. S. Abramov *et al.*, Landscape of allele-specific transcription factor binding in the human genome. *Nat. Commun.* **12**, 2751 (2021). doi: 10.1038/s41467-021-23007-0; pmid: 33980847
36. K. Blatt *et al.*, Identification of campath-1 (CD52) as novel drug target in neoplastic stem cells in 5q-patients with MDS and AML. *Clin. Cancer Res.* **20**, 3589–3602 (2014). doi: 10.1158/1078-0432.CCR-13-2811; pmid: 24799522
37. L. Chen *et al.*, Genetic drivers of epigenetic and transcriptional variation in human immune cells. *Cell* **167**, 1398–1414.e24 (2016). doi: 10.1016/j.cell.2016.10.026; pmid: 27863251
38. A. R. Martin *et al.*, Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584–591 (2019). doi: 10.1038/s41588-019-0379-x; pmid: 30926966
39. C.-Y. Chen *et al.*, Analysis across Taiwan Biobank, Biobank Japan and UK Biobank identifies hundreds of novel loci for 36 quantitative traits. medRxiv 2021.04.12.21255236 [Preprint] (2021); https://doi.org/10.1101/2021.04.12.21255236.
40. W. Zhou *et al.*, Global Biobank Meta-analysis Initiative: Powering genetic discovery across human disease. *Cell Genomics* **2**, 100192 (2022). doi: 10.1016/j.xgen.2022.100192; pmid: 36777996
41. M. M. Shull, J. B. Lingrel, Multiple genes encode the human Na+,K+-ATPase catalytic subunit. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 4039–4043 (1987). doi: 10.1073/pnas.84.12.4039; pmid: 3035563
42. N. Glorioso *et al.*, Association of ATP1A1 and dear single-nucleotide polymorphism haplotypes with essential hypertension: Sex-specific and haplotype-specific effects. *Circ. Res.* **100**, 1522–1529 (2007). doi: 10.1161/01.RES.0000267716.96196.60; pmid: 17446437
43. P. Araos, S. Figueroa, C. A. Amador, The role of neutrophils in hypertension. *Int. J. Mol. Sci.* **21**, 8536 (2020). doi: 10.3390/ijms21228536; pmid: 33198361
44. A. Buniello *et al.*, The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47** (D1), D1005–D1012 (2019). doi: 10.1093/nar/gky1120; pmid: 30445434
45. V. Leduc, S. Jasmin-Bélanger, J. Poirier, APOE and cholesterol homeostasis in Alzheimer's disease. *Trends Mol. Med.* **16**, 469–477 (2010). doi: 10.1016/j.molmed.2010.07.008; pmid: 20817608
46. S. Suchindran *et al.*, Genome-wide association study of Lp-PLA(2) activity and mass in the Framingham Heart Study. *PLOS Genet.* **6**, e1000928 (2010). doi: 10.1371/journal.pgen.1000928; pmid: 20442857
47. J. M. Replogle *et al.*, Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell* **185**, 2559–2575.e28 (2022). doi: 10.1016/j.cell.2022.05.013; pmid: 35688146
48. E. V. Fuior, A. V. Gafencu, Apolipoprotein C1: Its pleiotropic effects in lipid metabolism and beyond. *Int. J. Mol. Sci.* **20**, 5939 (2019). doi: 10.3390/ijms20235939; pmid: 31779116
49. A. Auton *et al.*, A global reference for human genetic variation. *Nature* **526**, 68–74 (2015). doi: 10.1038/nature15393; pmid: 26432245
50. M. P. Zafra *et al.*, Optimized base editors enable efficient editing in cells, organoids and mice. *Nat. Biotechnol.* **36**, 888–893 (2018). doi: 10.1038/nbt.4194; pmid: 29969439
51. R. T. Walton, K. A. Christie, M. N. Whittaker, B. P. Kleinstiver, Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* **368**, 290–296 (2020). doi: 10.1126/science.aba8853; pmid: 32217751
52. R. E. Hanna *et al.*, Massively parallel assessment of human variants with base editor screens. *Cell* **184**, 1064–1080.e20 (2021). doi: 10.1016/j.cell.2021.01.012; pmid: 33606977
53. R. Cuella-Martin *et al.*, Functional interrogation of DNA damage response variants with base editing screens. *Cell* **184**, 1081–1097.e19 (2021). doi: 10.1016/j.cell.2021.01.041; pmid: 33606978
54. N. Mancuso *et al.*, Integrating gene expression with summary association statistics to identify genes associated with 30 complex traits. *Am. J. Hum. Genet.* **100**, 473–487 (2017). doi: 10.1016/j.ajhg.2017.01.031; pmid: 28238358
55. S. H. Orkin, L. I. Zon, Hematopoiesis: An evolving paradigm for stem cell biology. *Cell* **132**, 631–644 (2008). doi: 10.1016/j.cell.2008.01.025; pmid: 18295580
56. J. J. Gasiorek, V. Blank, Regulation and function of the NFE2 transcription factor in hematopoietic and non-hematopoietic cells. *Cell. Mol. Life Sci.* **72**, 2323–2335 (2015). doi: 10.1007/s00018-015-1866-6; pmid: 25721735
57. K. L. Davis, Ikaros: Master of hematopoiesis, agent of leukemia. *Ther. Adv. Hematol.* **2**, 359–368 (2011). doi: 10.1177/2040620711412419; pmid: 23556102
58. R. J. Chan, R. Hromas, M. C. Yoder, The role of Hex in hemangioblast and hematopoietic development. *Methods Mol. Biol.* **330**, 123–134 (2006). doi: 10.1385/1-59745-036-7:123; pmid: 16846021
59. M. Ichikawa *et al.*, A role for RUNX1 in hematopoiesis and myeloid leukemia. *Int. J. Hematol.* **97**, 726–734 (2013). doi: 10.1007/s12185-013-1347-3; pmid: 23613270
60. E. Chapnik *et al.*, miR-142 orchestrates a network of actin cytoskeleton regulators during megakaryopoiesis. *eLife* **3**, e01964 (2014). doi: 10.7554/eLife.01964; pmid: 24859754
61. T. Wang, F. Wu, D. Yu, miR-144/451 in hematopoiesis and beyond. *ExRNA* **1**, 16 (2019). doi: 10.1186/s41544-019-0035-8
62. M. Ghanbari *et al.*, A functional variant in the miR-142 promoter modulating its expression and conferring risk of Alzheimer disease. *Hum. Mutat.* **40**, 2131–2145 (2019). doi: 10.1002/humu.23872; pmid: 31322790
63. C. Bellenguez *et al.*, New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat. Genet.* **54**, 412–436 (2022). doi: 10.1038/s41588-022-01024-z; pmid: 35379992
64. P. Shooshtarizadeh *et al.*, Gfi1b regulates the level of Wnt/β-catenin signaling in hematopoietic stem cells and megakaryocytes. *Nat. Commun.* **10**, 1270 (2019). doi: 10.1038/s41467-019-09273-z; pmid: 30894540

65. ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012). doi: 10.1038/nature11247; pmid: 22955616
66. V. Agarwal, G. W. Bell, J.-W. Nam, D. P. Bartel, Predicting effective microRNA target sites in mammalian mRNAs. *eLife* **4**, e05005 (2015). doi: 10.7554/eLife.05005; pmid: 26267216
67. S. E. McGeary *et al*., The biochemical basis of microRNA targeting efficacy. *Science* **366**, eaav1741 (2019). doi: 10.1126/science.aav1741; pmid: 31806698
68. G. R. Schnitzler *et al*., Mapping the convergence of genes for coronary artery disease onto endothelial cell programs. bioRxiv 514606 [Preprint] (2022); https://doi.org/10.1101/2022.11.01.514606.
69. A. Regev *et al*., Science Forum: The Human Cell Atlas. *eLife* **6**, e27041 (2017). doi: 10.7554/eLife.27041; pmid: 29206104
70. T. Stuart *et al*., Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019). doi: 10.1016/j.cell.2019.05.031; pmid: 31178118
71. T. Möröy, L. Vassen, B. Wilkes, C. Khandanpour, From cytopenia to leukemia: The role of Gfi1 and Gfi1b in blood formation. *Blood* **126**, 2561–2569 (2015). doi: 10.1182/blood-2015-06-655043; pmid: 26447191
72. E. Anguita, F. J. Candel, A. Chaparro, J. J. Roldán-Etcheverry, Transcription factor GFI1B in health and disease. *Front. Oncol.* **7**, 54 (2017). doi: 10.3389/fonc.2017.00054; pmid: 28401061
73. H. Beauchemin, T. Möröy, Multifaceted actions of GFI1 and GFI1B in hematopoietic stem cell self-renewal and lineage commitment. *Front. Genet.* **11**, 591099 (2020). doi: 10.3389/fgene.2020.591099; pmid: 33193732
74. M. Osawa *et al*., Erythroid expansion mediated by the Gfi-1B zinc finger protein: Role in normal hematopoiesis. *Blood* **100**, 2769–2777 (2002). doi: 10.1182/blood-2002-01-0182; pmid: 12351384
75. F. Garnache-Ottou *et al*., Expression of the myeloid-associated marker CD33 is not an exclusive factor for leukemic plasmacytoid dendritic cells. *Blood* **105**, 1256–1264 (2005). doi: 10.1182/blood-2004-06-2416; pmid: 15388576
76. M. S. De Propris *et al*., High CD33 expression levels in acute myeloid leukemia cells carrying the nucleophosmin (NPM1) mutation. *Haematologica* **96**, 1548–1551 (2011). doi: 10.3324/haematol.2011.043786; pmid: 21791474
77. A. Dixit *et al*., Perturb-Seq: Dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* **167**, 1853–1866.e17 (2016). doi: 10.1016/j.cell.2016.11.038; pmid: 27984732
78. B. Adamson *et al*., A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell* **167**, 1867–1882.e21 (2016). doi: 10.1016/j.cell.2016.11.048; pmid: 27984733
79. S.-Y. Yu *et al*., Increasing the targeting scope of CRISPR base editing system beyond NGG. *CRISPR J.* **5**, 187–202 (2022). doi: 10.1089/crispr.2021.0109; pmid: 35238621
80. Z. Daniloski *et al*., Identification of required host factors for SARS-CoV-2 infection in human cells. *Cell* **184**, 92–105.e16 (2021). doi: 10.1016/j.cell.2020.10.030; pmid: 33147445
81. N. Liscovitch-Brauer *et al*., Profiling the genetic determinants of chromatin accessibility with scalable single-cell CRISPR screens. *Nat. Biotechnol.* **39**, 1270–1277 (2021). doi: 10.1038/s41587-021-00902-x; pmid: 33927415
82. M. Legut *et al*., A genome-scale screen for synthetic drivers of T cell proliferation. *Nature* **603**, 728–735 (2022). doi: 10.1038/s41586-022-04494-7; pmid: 35296855
83. C. Benner *et al*., FINEMAP: Efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493–1501 (2016). doi: 10.1093/bioinformatics/btw018; pmid: 26773131
84. S. Purcell *et al*., PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007). doi: 10.1086/519795; pmid: 17701901
85. G. Band, J. Marchini, BGEN: a binary file format for imputed genotype and haplotype data. bioRxiv 308296 (2018).
86. C. Benner *et al*., Prospects of fine-mapping trait-associated genomic regions by using summary statistics from genome-wide association studies. *Am. J. Hum. Genet.* **101**, 539–551 (2017). doi: 10.1016/j.ajhg.2017.08.012; pmid: 28942963
87. W. McLaren *et al*., The Ensembl variant effect predictor. *Genome Biol.* **17**, 122 (2016). doi: 10.1186/s13059-016-0974-4; pmid: 27268795
88. J. B. Maller *et al*., Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat. Genet.* **44**, 1294–1301 (2012). doi: 10.1038/ng.2435; pmid: 23104008
89. R. Mägi *et al*., Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Hum. Mol. Genet.* **26**, 3639–3650 (2017). doi: 10.1093/hmg/ddx280; pmid: 28911207
90. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010). doi: 10.1093/bioinformatics/btq033; pmid: 20110278
91. S. Neph *et al*., BEDOPS: High-performance genomic feature operations. *Bioinformatics* **28**, 1919–1920 (2012). doi: 10.1093/bioinformatics/bts277; pmid: 22576172
92. M. Legut *et al*., High-throughput screens of PAM-flexible Cas9 variants for gene knockout and transcriptional modulation. *Cell Rep.* **30**, 2859–2868.e5 (2020). doi: 10.1016/j.celrep.2020.02.010; pmid: 32130891
93. B. Chen *et al*., Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479–1491 (2013). doi: 10.1016/j.cell.2013.12.001; pmid: 24360272
94. H.-H. Wessels *et al*., Massively parallel Cas13 screens reveal principles for guide RNA design. *Nat. Biotechnol.* **38**, 722–727 (2020). doi: 10.1038/s41587-020-0456-9; pmid: 32518401
95. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009). doi: 10.1186/gb-2009-10-3-r25; pmid: 19261174
96. R. Kolde, S. Laur, P. Adler, J. Vilo, Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics* **28**, 573–580 (2012). doi: 10.1093/bioinformatics/btr709; pmid: 22247279
97. R. M. Meyers *et al*., Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat. Genet.* **49**, 1779–1784 (2017). doi: 10.1038/ng.3984; pmid: 29083409
98. J. M. Dempster *et al*., Extracting biological insights from the Project Achilles genome-scale CRISPR screens in cancer cell lines. bioRxiv 720243 [Preprint] (2019); https://doi.org/10.1101/720243.
99. J. M. Dempster *et al*., Chronos: a CRISPR cell population dynamics model. bioRxiv 432728 [Preprint] (2021); https://doi.org/10.1101/2021.02.25.432728.
100. C. Pacini *et al*., Integrated cross-study datasets of genetic dependencies in cancer. *Nat. Commun.* **12**, 1661 (2021). doi: 10.1038/s41467-021-21898-7; pmid: 33712601
101. M. G. Kluesner *et al*., CRISPR-Cas9 cytidine and adenosine base editing of splice-sites mediates highly-efficient disruption of proteins in primary and immortalized cells. *Nat. Commun.* **12**, 2437 (2021). doi: 10.1038/s41467-021-22009-2; pmid: 33893286
102. W. F. Lenoir, T. L. Lim, T. Hart, PICKLES: The database of pooled in-vitro CRISPR knockout library essentiality screens. *Nucleic Acids Res.* **46** (D1), D776–D780 (2018). doi: 10.1093/nar/gkx993; pmid: 29077937
103. A. C. Komor, Y. B. Kim, M. S. Packer, J. A. Zuris, D. R. Liu, Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016). doi: 10.1038/nature17946; pmid: 27096365
104. G. X. Y. Zheng *et al*., Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017). doi: 10.1038/ncomms14049; pmid: 28091601
105. Y. Hao *et al*., Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021). doi: 10.1016/j.cell.2021.04.048; pmid: 34062119
106. K. Alasoo *et al*., Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. *Nat. Genet.* **50**, 424–431 (2018). doi: 10.1038/s41588-018-0046-7; pmid: 29379200
107. Y. Nédélec *et al*., Genetic ancestry and natural selection drive population differences in immune responses to pathogens. *Cell* **167**, 657–669.e21 (2016). doi: 10.1016/j.cell.2016.09.025; pmid: 27768889
108. H. Quach *et al*., Genetic adaptation and Neandertal admixture shaped the immune system of human populations. *Cell* **167**, 643–656.e17 (2016). doi: 10.1016/j.cell.2016.09.024; pmid: 27768888
109. B. J. Schmiedel *et al*., Impact of genetic polymorphisms on human immune cell gene expression. *Cell* **175**, 1701–1715.e16 (2018). doi: 10.1016/j.cell.2018.10.022; pmid: 30449622
110. A. Buil *et al*., Gene-gene and gene-environment interactions detected by transcriptome sequence analysis in twins. *Nat. Genet.* **47**, 88–91 (2015). doi: 10.1038/ng.3162; pmid: 25436857
111. T. Lappalainen *et al*., Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**, 506–511 (2013). doi: 10.1038/nature12531; pmid: 24037378
112. M. Gutierrez-Arcelus *et al*., Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *eLife* **2**, e00523 (2013). doi: 10.7554/eLife.00523; pmid: 23755361
113. E. Theusch, Y. I. Chen, J. I. Rotter, R. M. Krauss, M. W. Medina, Genetic variants modulate gene expression statin response in human lymphoblastoid cell lines. *BMC Genomics* **21**, 555 (2020). doi: 10.1186/s12864-020-06966-4; pmid: 32787775
114. F. Aguet *et al*., The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020). doi: 10.1126/science.aaz1776; pmid: 32913098
115. K. Lepik *et al*., C-reactive protein upregulates the whole blood expression of CD59 - an integrative analysis. *PLOS Comput. Biol.* **13**, e1005766 (2017). doi: 10.1371/journal.pcbi.1005766; pmid: 28922377
116. H. Kilpinen *et al*., Common genetic variation drives molecular heterogeneity in human iPSCs. *Nature* **546**, 370–375 (2017). doi: 10.1038/nature22403; pmid: 28489815
117. A. D. Panopoulos *et al*., iPSCORE: A resource of 222 iPSC lines enabling functional characterization of genetic variation across a variety of cell types. *Stem Cell Reports* **8**, 1086–1100 (2017). doi: 10.1016/j.stemcr.2017.03.012; pmid: 28410642
118. E. E. Pashos *et al*., Large, diverse population cohorts of hiPSCs and derived hepatocyte-like cells reveal functional genetic variation at blood lipid-associated loci. *Cell Stem Cell* **20**, 558–570.e10 (2017). doi: 10.1016/j.stem.2017.03.017; pmid: 28388432
119. B. E. Gryder, J. Khan, B. Z. Stanton, Measurement of differential chromatin interactions with absolute quantification of architecture (AQuA-HiChIP). *Nat. Protoc.* **15**, 1209–1236 (2020). doi: 10.1038/s41596-019-0285-9; pmid: 32051612
120. N. Servant *et al*., HiC-Pro: An optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015). doi: 10.1186/s13059-015-0831-x; pmid: 26619908
121. S. Bhattacharyya, V. Chandra, P. Vijayanand, F. Ay, Identification of significant chromatin contacts from HiChIP data by FitHiChIP. *Nat. Commun.* **10**, 4221 (2019). doi: 10.1038/s41467-019-11950-y; pmid: 31530818
122. D. van Dijk *et al*., Recovering gene interactions from single-cell data using data diffusion. *Cell* **174**, 716–729.e27 (2018). doi: 10.1016/j.cell.2018.05.061; pmid: 29961576
123. P. Langfelder, S. Horvath, WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008). doi: 10.1186/1471-2105-9-559; pmid: 19114008
124. J. A. Griffiths, A. C. Richard, K. Bach, A. T. L. Lun, J. C. Marioni, Detection and removal of barcode swapping in single-cell RNA-seq data. *Nat. Commun.* **9**, 2667 (2018). doi: 10.1038/s41467-018-05083-x; pmid: 29991676
125. A. T. L. Lun *et al*., EmptyDrops: Distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biol.* **20**, 63 (2019). doi: 10.1186/s13059-019-1662-y; pmid: 30902100

ACKNOWLEDGMENTS

We thank the staff of the Sanjana and Lappalainen laboratories for support and advice; the New York Genome Center for flow cytometry resources; E. Hoelzli for support with digital PCR; T. Stuart and H. Wessels for advice on single-cell sequencing quality control; and S. Kasela for help with library design. This research was conducted using the UKBB Resource under application no. 47976. **Funding:** J.A.M. is supported by a Canadian Institutes of Health Research Banting Postdoctoral Fellowship and the National Institutes of Health (NIH)/National Human Genome Research Institute (NHGRI) (grant K99HG012792). Z.D. is supported by an American Heart Association Postdoctoral Fellowship (20POST35220040). J.D. is supported by a European Molecular Biology Organization Postdoctoral Fellowship (ALTF 345-2021). K.R. is supported by the NIH/National Institute of Mental Health (NIMH) (grant R01MH123184). E.K. is supported by the National Science Foundation (NSF) (grant DMS-2113072) and the Wharton

Data Science and Business Analytics Fund. T.L. is supported by the NIH/NIMH (grant R01MH106842), the NIH/NHGRI (grant UM1HG008901), and the NIH/National Institute of General Medical Sciences (NIGMS) (grant R01GM122924). N.E.S. is supported by the NIH/NHGRI (grant DP2HG010099, including a supplemental award for STING-seq), the NIH/National Cancer Institute (NCI) (grants R01CA279135 and R01CA218668), the NIH/National Institute of Allergy and Infectious Diseases (NIAID) (grant R01AI176601), the Simons Foundation for Autism Research, the MacMillan Center for the Study of the Non-Coding Cancer Genome, New York University, and New York Genome Center funds. **Author contributions:** J.A.M. and N.E.S. conceived the project. J.A.M., D.A.G., and N.E.S. performed GWAS statistical fine-mapping, functional annotation, and STING-seq library design. J.A.M., Z.D., M.Z., L.L., and K.D. developed the lentiviral constructs and monoclonal K562 cells. J.A.M., Z.D., and L.L. designed and performed lentiviral transductions. C.C. performed the essential gene pooled screens and physical contact mapping experiments. J.A.M., S.H., and E.P.M. performed single-cell sequencing. J.A.M., T.B., and E.K. performed differential gene expression– and protein-level testing. J.A.M. and J.D. performed coexpression analyses and human bone marrow single-cell data analyses. P.S., K.R., E.K., T.L., and N.E.S. supervised the work. J.A.M., T.L., and N.E.S. wrote the manuscript with input from all authors. **Competing interests:** N.E.S. is an adviser to Qiagen and is a cofounder and adviser of OverT Bio. T.L. is an adviser to Goldfinch Bio and GSK and, with equity, Variant Bio. The remaining authors declare no competing interests. **Data and materials availability:** Plasmids (lentiCRISPRi(v1)-Blast, lentiCRISPRi(v2)-Blast, lentiGuideFE-Puro, and lentiBE3-SpRY-Blast) have been deposited with Addgene (plasmid nos. 170067, 170068, 170069, 199303). Raw and processed single-cell sequencing files are available from Gene Expression Omnibus (GEO) under accession number GSE171452. SCEPTRE and related tutorials are available on Github (www.github.com/Katsevich-Lab/sceptre). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adh7699
Figs. S1 to S17
Tables S1 to S4
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 13 March 2023; accepted 20 April 2023
Published online 4 May 2023
10.1126/science.adh7699

RESEARCH ARTICLE

ASYMMETRIC CATALYSIS

Generality-oriented optimization of enantioselective aminoxyl radical catalysis

Jonas Rein[1]†, Soren D. Rozema[2]†, Olivia C. Langner[2]‡, Samson B. Zacate[1]‡, Melissa A. Hardy[3], Juno C. Siu[1], Brandon Q. Mercado[2], Matthew S. Sigman[3]*, Scott J. Miller[2]*, Song Lin[1]*

Catalytic enantioselective methods that are generally applicable to a broad range of substrates are rare. We report a strategy for the oxidative desymmetrization of *meso*-diols predicated on a nontraditional catalyst optimization protocol by using a panel of screening substrates rather than a singular model substrate. Critical to this approach was rational modulation of a peptide sequence in the catalyst incorporating a distinct aminoxyl-based active residue. A general catalyst emerged, providing high selectivity in the delivery of enantioenriched lactones across a broad range of diols, while also achieving up to ~100,000 turnovers.

Substrate generality is a long-standing aim in the field of enantioselective catalysis (*1*–*3*). Even so, specificity-oriented optimization has been broadly adopted in the field, resulting in many catalysts that are tailored for one model substrate but often exhibit diminished selectivity upon surveying a diverse substrate scope (Fig. 1A) (*4*). To address the limited transferability frequently encountered in asymmetric catalysis, we envisioned an optimization strategy with generality as the primary objective in catalyst development. This strategy requires parallel screening of a large, diverse catalyst library against a judiciously selected panel of model substrates that represent the chemical space of the target substrate class, rather than a singular substrate. Such an approach is reminiscent of the "one-pot multi-substrate" method to increase screening throughput (*5*–*8*). However, this pooled-substrate method introduces challenges in chemical compatibility and product analysis and thus has seen limited implementation in catalyst development.

Recently, Jacobsen and co-workers reported a pooled-product method for enantioselectivity analysis, which enabled high-throughput evaluation of the generality of known catalysts in the context of an asymmetric Pictet-Spengler reaction (*4*). Our interest focused on advancing the notion of enhanced substrate breadth in the context of previously elusive transformations by using a new catalytic platform. Toward this objective, we envisioned that small synthetic, catalytic peptides would be well suited for such a study, given the modularity of their constituent amino acids and their synthetic accessibility, which provides a strong foundation for a generality-driven optimization campaign (*9*). Peptide optimization can also bear a resemblance to the venerable directed evolution of enzymes through survey of multidimensional catalyst space (*10*, *11*). Our study combined the vast amount of catalyst space surveyed with a computationally informed representative substrate space, yielding information-rich data that can be used for subsequent catalyst optimization. We investigated this strategy in an underexplored area in asymmetric synthesis: enantioselective aminoxyl radical catalysis (*12*).

Aminoxyl radicals are a class of persistent open-shell molecules that have found broad applications in organic synthesis (*13*). They are excellent catalysts for the oxidation of various functional groups such as alcohols, amines, and alkenes through intermediate oxoammonium ions (*13*–*18*). In contrast to the vast number of reactions known to be catalyzed by aminoxyl radicals, enantioselective variants remain rare, hampered by a dearth of synthetically accessible and tunable catalysts. Foundational work by the group of Bobbitt and recent advances by Iwabuchi and co-workers among others provided salient precedents in asymmetric alcohol oxidation (Fig. 1B) (*19*–*22*). Nevertheless, these catalysts require lengthy synthetic preparation and lack the structural modularity needed for generality-oriented optimization; accordingly, substrate scope for these efforts suffers from diminished enantioselectivities beyond a handful of similar substrates. Toniolo and co-workers demonstrated that oxoammonium ions are compatible within peptidic frameworks and observed the kinetic resolution (KR) of a singular alcohol substrate with modest selectivities [relative rate constant (k_{rel}) < 3] (*12*).

We hypothesized that incorporation of an achiral aminoxyl moiety into diverse peptide-based scaffolds would enable rapid substrate exploration within the reaction space accessible to chiral aminoxyl catalysis. Accordingly, we designed and synthesized an aminoxyl monomer (**Azc-OMe**) (Fig. 1C); the azaadamantyl aminoxyl framework was selected for its excellent activity in catalytic alcohol oxidation resulting from its high oxidation potential, enhanced steric accessibility compared with that of monocyclic aminoxyls, and conformational rigidity (*23*). Additionally, a methyl ester adjacent to the aminoxyl residue enables rapid incorporation into peptide frameworks with the aminoxyl moiety proximal to the chiral functionality. This monomer design allowed for the synthesis of >70 chiral aminoxyl peptides. We report the development of the enantioselective desymmetrization of *meso*-diols, for which a generality-oriented optimization campaign led to an aminoxyl-based oligopeptide catalyst that delivers high selectivity for accessing structurally diverse enantioenriched lactones.

Initial studies using a traditional optimization strategy

We began our study with the desymmetrization of *meso*-diols (Fig. 2A), which proceeds with an initial oxidation of one of the enantiotopic hydroxyl groups, followed by cyclization through attack of the second hydroxyl group on the incipient aldehyde to furnish an intermediate chiral lactol (**H**). A second oxidation then delivers a lactone (**L**). Chiral lactones are valuable and versatile building blocks in organic and polymer synthesis as well as medicinal chemistry. Although methods are available for synthesizing specific types of enantioenriched lactones, a broadly applicable approach that is compatible with a wide range of substrate structures, ring sizes, and substitution patterns remains elusive and desirable. In particular, few reports achieved chiral lactone synthesis through the oxidative desymmetrization of readily available *meso*-diols; these methods—which use either a transition metal (iridium or ruthenium) (*24*, *25*) or organocatalysis (*19*, *26*–*28*)—show limited substrate diversity and/or moderate enantioselectivity.

We initially used a traditional optimization approach by screening a collection of aminoxyl-embedded peptide catalysts in the oxidation of 1,4-diol **S1**, using trichloroisocyanuric acid (TCCA) as a terminal oxidant and sodium bicarbonate as a base (*20*, *21*). This effort yielded peptide **P1**, which provided lactone **L1** with substantial enantioselectivity [75% enantiomeric excess (*ee*)]. However, **P1** showed diminished selectivity when tested against an array of additional 1,4-diols that display a variety of steric and electronic profiles, giving a median ee (ee_{med}) of only 16% (Fig. 2E, column **P1**). This

[1]Department of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853, USA. [2]Department of Chemistry, Yale University, New Haven, CT 06520, USA. [3]Department of Chemistry, University of Utah, Salt Lake City, UT 84112, USA.
*Corresponding author. Email: songlin@cornell.edu (S.L.); scott.miller@yale.edu (S.J.M.); matt.sigman@utah.edu (M.S.S.)
†These authors contributed equally to this work.
‡These authors contributed equally to this work.

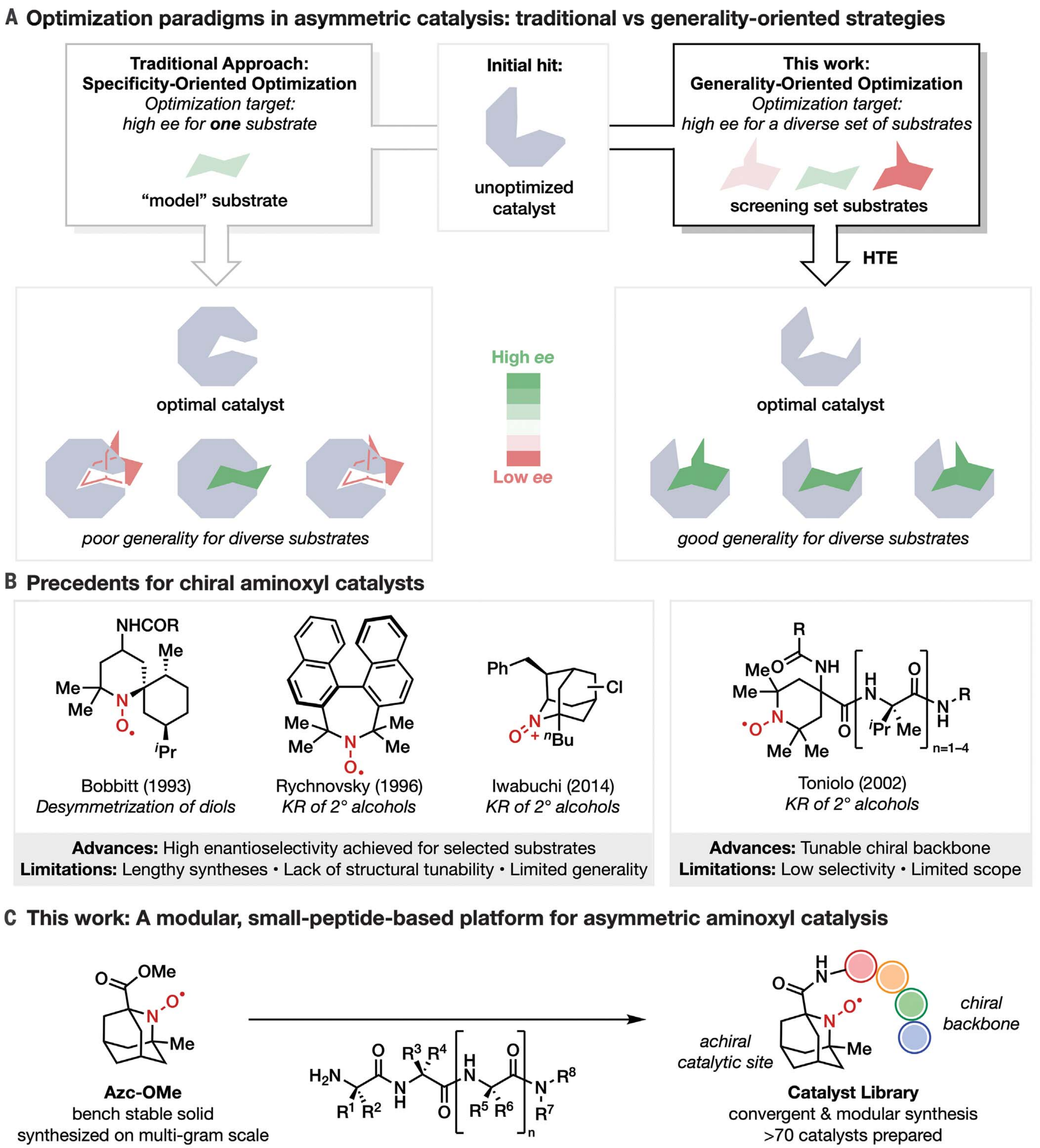
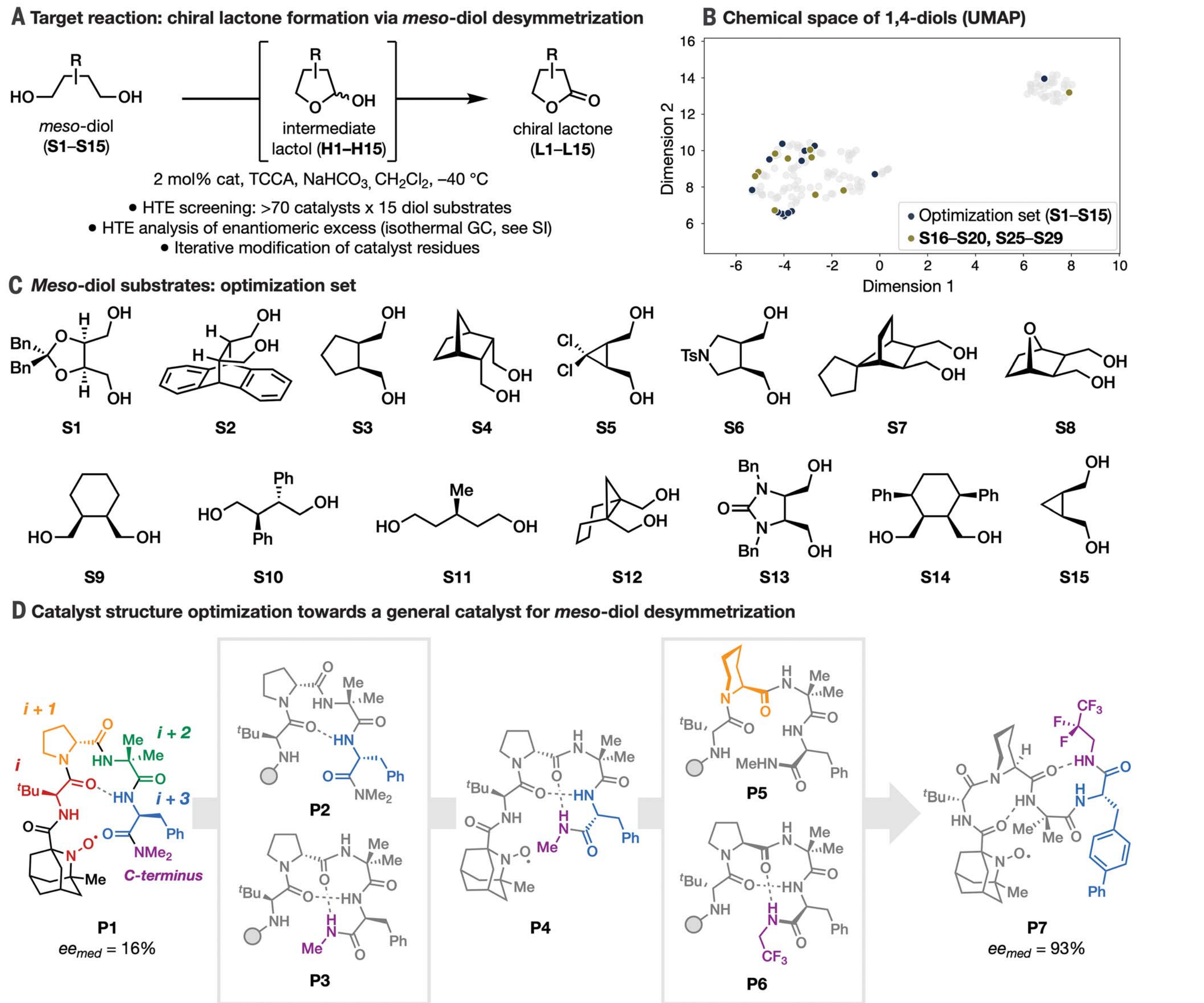


Fig. 1. A generality-oriented optimization strategy for enantioselective catalysis with peptidyl aminoxyl radicals. (**A**) Optimization paradigms in asymmetric catalysis: traditional versus generality-oriented strategies. (**B**) Precedents for chiral aminoxyl and oxoammonium catalysts (*12*, *19*, *20*, *22*). (**C**) This work: A modular, small-peptide-based platform for aminoxyl asymmetric catalysis. **Azc-OMe** was accessed in seven steps with a single chromatographic separation.

result underscores the challenge of identifying highly selective catalysts that tolerate a broad scope when focusing on a singular substrate during catalyst optimization.

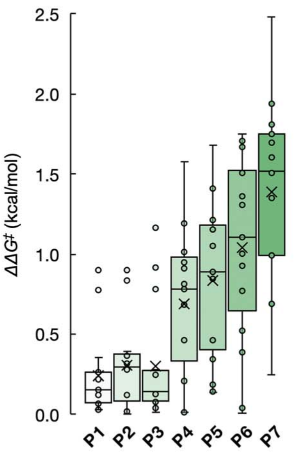
Generality-oriented catalyst optimization guided by median *ee*

Aiming for a more broadly applicable method, we assessed a diverse library of peptide scaffolds with high-throughput experimentation (HTE) using 24-well plate parallel screening and gas chromatography analysis (supplementary materials). Using this workflow, we discuss



E Optimization data

%ee	P1	P2	P3	P4	P5	P6	P7
S1	75	75	85	94	95	95	97
S2	6	2	11	69	83	86	97
S3	19	30	26	75	72	93	96
S4	23	33	1	77	91	90	96
S5	68	72	76	86	86	95	95
S6	36	37	69	87	85	96	94
S7	27	40	4	63	82	83	93
S8	2	13	14	71	74	89	93
S9	8	0	29	69	48	79	93
S10	27	40	28	48	81	68	90
S11	3	2	8	46	36	76	79
S12	7	5	17	22	46	51	79
S13	13	29	6	4	15	40	65
S14	9	34	15	1	15	4	63
S15	16	31	9	1	20	1	26
%ee_{med}	16	31	15	69	74	83	93



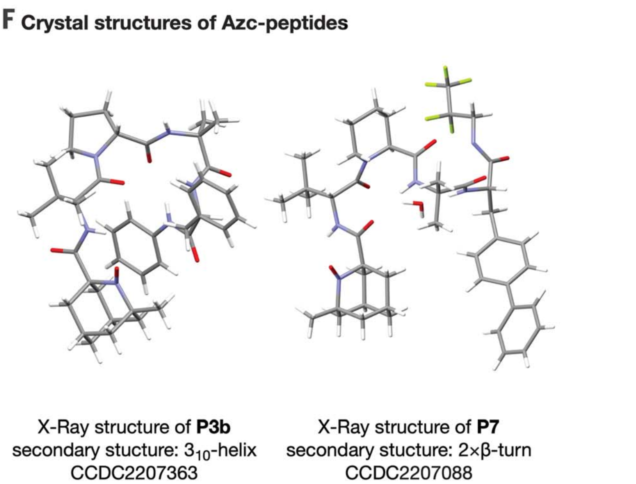


Fig. 2. Generality-oriented optimization through combinatorial screening of catalysts with a substrate library and iterative modification of catalyst structure. (**A**) Chiral lactone formation through *meso*-diol desymmetrization. (**B**) Chemical space of *meso*-1,4-diols defined by using Mordred descriptors and a uniform manifold approximation and projection (UMAP) with molecular fingerprints of commercial *meso*-1,4-diols from a Reaxys search. (**C**) *meso*-Diol substrates: optimization set. (**D**) Catalyst structure optimization. (**E**) Optimization data and box and whisker chart for the calculated $\Delta\Delta G^{\ddagger}$ for the optimization data. (**F**) Crystal structures of peptides **P3b** and **P7**.

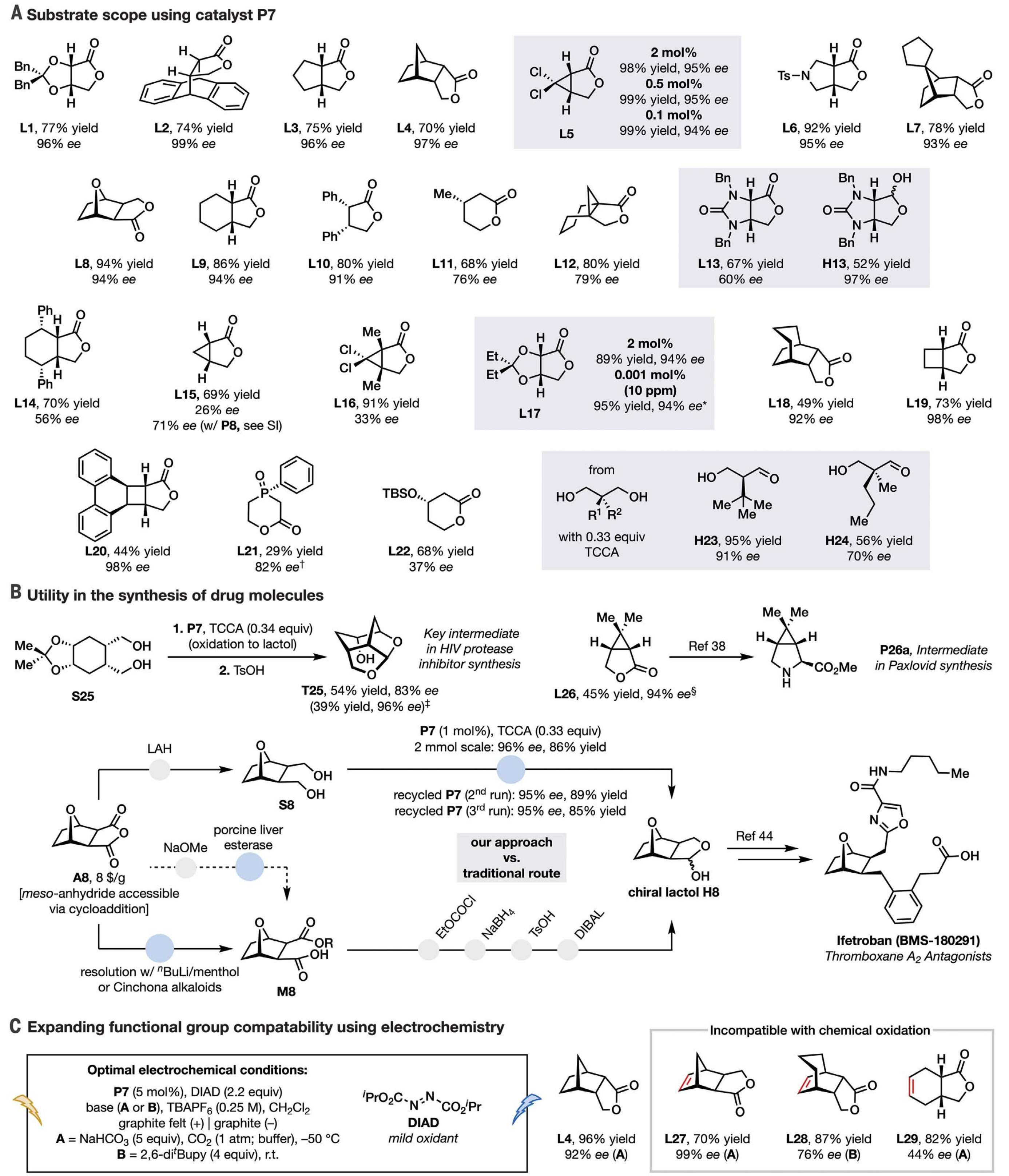


Fig. 3. Substrate scope. (**A**) Substrate scope by using catalyst **P7**. (**B**) Utility in the synthesis of drug molecules. (**C**) Expanding functional group compatibility by using electrochemistry. Absolute stereochemistry was assigned by analogy to **L1**, **L4**, and **L6**. *0.001 mol % **P7**, 96 hours, −50°C, then 2 mol % ACT, 0.3 equiv TCCA, room temperature (r.t.), 6 hours. †Isolated as the methyl ester after SiO_2 catalyzed methanolysis. ‡After trituration with pentane; §10 mol % **P7**.

the data obtained with a given catalyst in terms of the ee_{med}, which served as the primary optimization objective; excellent conversion and clean reaction profiles for each substrate enabled a narrow focus of the optimization on enantioselectivity. Substrate selection was guided by mapping the chemical space of commercially available *meso*-1,4-diols to ensure diversity of the screening set (Fig. 2B). This was accomplished by computing molecular descriptors of all commercially available *meso*-diols and applying dimensionality reduction techniques (supplementary materials) (*29*, *30*). This analysis revealed two clusters of diols, with the larger cluster representing disubstituted *meso*-diols, and the smaller cluster consisting of a distinct group of 2,2,3,3-tetrasubstitued 1,4-diols. From this collection of compounds, 14 structurally diverse substrates with a variety of functionalities were included in the optimization set. On the basis of their relative position within the uniform manifold approximation and projection (UMAP) and synthetic accessibility, we settled on eight monocyclic (**S1**, **S3**, **S5**, **S6**, **S9**, and **S13** to **15**) and five polycyclic (**S2**, **S4**, **S7**, **S8**, and **S12**) substrates with various ring sizes, as well as one acyclic diol (**S10**), to function as our representative set of diols (Fig. 2C). Additionally, we included a 1,5-diol (**S11**) to further increase diversity in the optimization set.

With this set of model substrates, we then evaluated catalysts as a function of their constituent residues (Fig. 2D). Beginning with **P1** (ee_{med} = 16%), two single-point changes exhibited increases in *ee*: inversion of stereochemistry at *i+3* (**P2**) provided ee_{med} of 31%, and replacement of the NMe_2 group at the C terminus with NHMe (**P3**) provided an ee_{med} of 15% but displayed considerable improvement for three substrates (**S1**, **S6**, **S9**). Auspiciously, **P3b**, an analog of **P3** maintaining the key C-terminal secondary amide, was crystalline, allowing determination of the secondary structure by means of single-crystal x-ray diffraction analysis (Fig. 2F); rotating-frame nuclear Overhauser effect correlation spectroscopy (ROESY) nuclear magnetic resonance (NMR) confirmed that the peptide adopts a 3_{10}-helical motif in solution (*31*). The confluence of changes leading to **P2** and **P3** from **P1** was found to be beneficial; **P4**, possessing both the optimized relative stereochemical configuration of **P2** and the NHMe C terminus of **P3**, provided a considerable enhancement in selectivity, with an ee_{med} of 69% (**P4**). Further modifications to **P4** showed that two additional point changes furnished substantial enhancement in ee_{med}: Replacement of the five-membered proline (Pro) with the six-membered homolog pipecolic acid (Pip) at *i* + 1 (**P5**) led to an increase of ee_{med} to 74%; concurrently, the inclusion of the fluorinated $NHCH_2CF_3$ substituent at the C-terminal

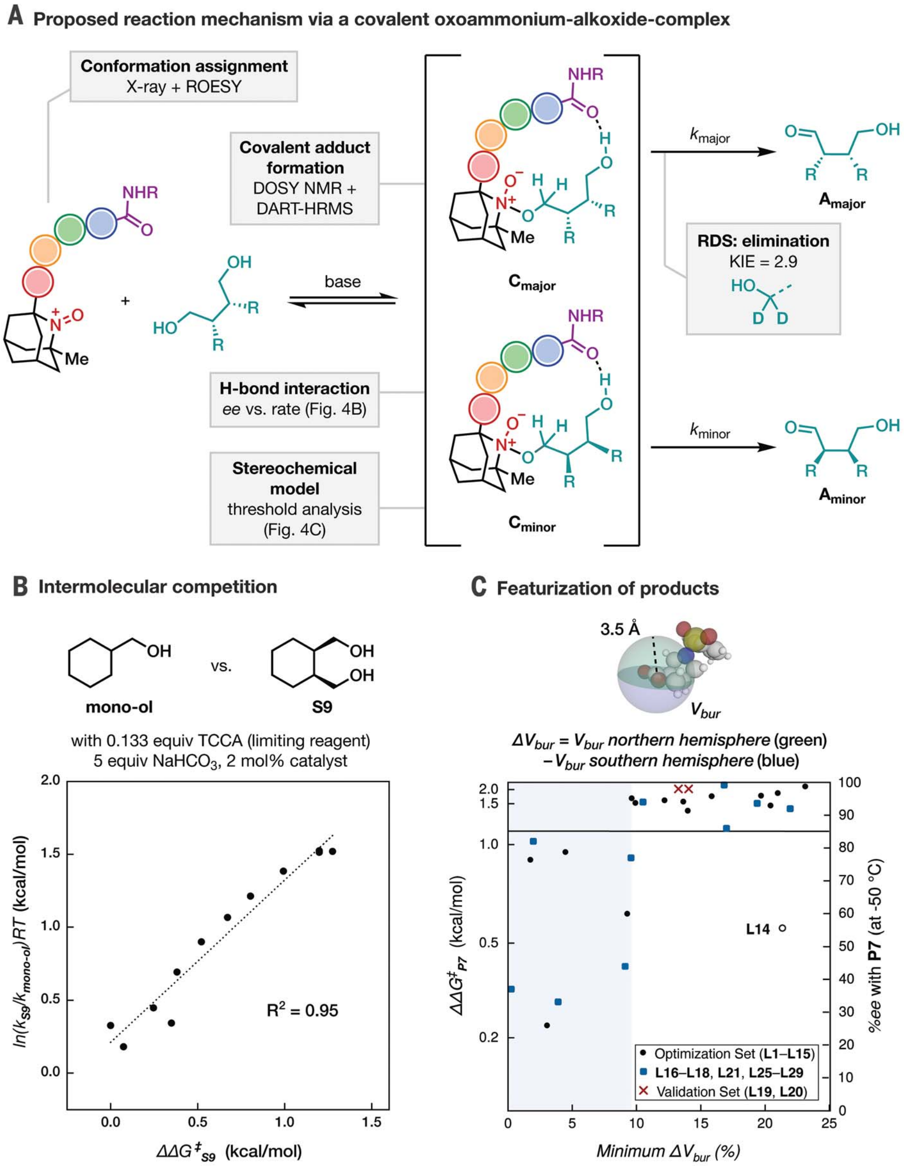


Fig. 4. Mechanistic analysis. (**A**) Proposed reaction mechanism through a covalent oxoammonium-alkoxide complex. (**B**) Intermolecular competition studies. (**C**) Featurization of products (1,3-diols **S23** and **S24** were omitted from the analysis because they do not form lactones upon oxidation). A discussion of **L14** as a false positive is available in the supplementary materials.

position resulted in peptide **P6**, which greatly outperformed catalysts with other C termini, exhibiting an ee_{med} of 83%. Catalyst **P7**, which combined the Pip residue found in **P5** with an elongated polyfluorinated C terminus ($NHCH_2C_2F_5$) and a change to biphenylalanine (Bip) at *i* + 3, gave a further boost of ee_{med} to 93%, with 10 of 15 substrates showing >90% *ee*, and an additional four substrates displaying *ee* > 60%. Single-crystal x-ray diffraction analysis and ROESY NMR were also performed on **P7**, revealing that both the aminoxyl radical and the active oxoammonium forms of the catalyst share secondary structural features. The solid-state structure of **P7** shows a departure from the 3_{10}-helix observed in **P3b**, displaying instead two successive β-turns (Fig. 2F) (*32*). Because it appears that multiple secondary structures can contribute to effective enantioinduction, attributing high enantioselectivity to a singular stereochemical model may not reflect the behavior of an ensemble. Several conformationally distinct transition states are plausible—an emerging feature of flexible, highly effective enantioselective catalysts (*32*, *33*) and a likely basis for the observed high selectivity across a diverse collection of *meso*-diols.

Substrate scope and synthetic applications

We then carried out the desymmetrization on a synthetically relevant scale (0.2 mmol) using the optimal catalyst **P7** (Fig. 3A). In addition to the 15 original substrates selected for the optimization campaign, we included a panel of *meso*-diols with varying chain lengths (1,3-, 1,4-, and 1,5-diols) and skeletal arrangements that further sample the chemical space (Fig. 2B and 4C) to define the scope and limitations of this method. We found that ketal-protected erythronic acid γ-lactones **L1** and **L17** were formed in good yields and excellent enantioenrichment. Polycyclic diols were efficiently desymmetrized (**L2**, **L4**, **L7**, **L8**, and **L18**), providing valuable precursors for pharmaceuticals and chiral ligands (*34*–*36*). A suite of fused bicyclic lactones was also efficiently synthesized. For example, cyclopentane lactone **L3** and fungistatic cyclohexane lactone **L9** (*37*) were formed with 96% *ee* and 94% *ee*, respectively, and cyclobutane-derived diols provided the corresponding [5.4.0]-fused bicyclic lactones **L19** and **L20** in 98% *ee*. Evaluation of cyclopropane diols revealed a requirement for substitution on the ring for high selectivity; unsubstituted **L15** was accessed in only 26% *ee*; however, dichloro- and dimethylcyclopropane lactones **L5** and **L26** were produced with excellent enantioselectivity, the latter of which is a key intermediate in the preparation of numerous important targets, including COVID drug nirmatrelvir (Paxlovid) (Fig. 3B) (*38*, *39*). Parenthetically, a structurally distinct catalyst **P8** (supplementary materials, section 4) was identified that provided **L15** in 71% *ee*, further highlighting the advantage of the modular catalyst screening platform. Pyrrolidine-bearing **S6** was desymmetrized with excellent yield (92%) and selectivity (95% *ee*). Skeletally distinct linear 1,4-diols were also good substrates in this reaction; lactone **L10** was obtained in 80% yield and 91% *ee*. Tetrasubstituted diols were also compatible, with bicyclic substrate **S12** yielding a [4.3.1]-propellane that featured two contiguous quaternary stereogenic centers in 80% yield and 79% *ee*; tetrasubstituted lactone **L16**, however, was obtained in only 33% *ee*.

To further evaluate the limit of our approach, we performed the desymmetrization of prochiral linear 1,5-diol and 1,3-diols. Diol **S11**, bearing a methyl group at the central carbon, was transformed to δ-lactone **L11** with 76% *ee*, providing access to a monomer for a biodegradable isotactic polymer (*40*). The catalyst was also competent in desymmetrization of a prochiral-at-phosphorous 1,5-diol, affording lactone **L21** in 82% *ee*. An additional 1,5-diol exhibited lower selectivity, with **L22** being enriched to only 37% *ee*. Notably, 1,3-diols, which were completely absent from the substrate screening set, also provided high enantioselectivities, yielding mono-oxidized β-hydroxy aldehydes as valuable synthetic precursors. 2-*tert*-butyl diol **S23** was efficiently desymmetrized in both excellent yield (95%) and enantioselectivity (91% *ee*). A disubstituted 1,3-diol **S24** yielded an α-quaternary β-hydroxy aldehyde (**H24**), a challenging motif to access in high enantioenrichment through alternative methods, differentiating methyl and propyl groups in 70% *ee*.

In the above cases, the reported enantiomeric excess reflects the selectivity in the initial desymmetrization step rather than that of the subsequent lactol oxidation because (i) for most substrates, we observed minimal stereochemical control of the second oxidation by the catalyst, and (ii) we carried out the reaction with an excess of TCCA to ensure full conversion of the intermediate lactol to the lactone product. However, we observed in two cases unusual kinetic resolution in either of the two oxidation steps, which was leveraged to improve enantioselectivity by decreasing the loading of TCCA. For example, diphenylcyclohexyl diol **S14**, which throughout optimization gave poor selectivities, was desymmetrized up to 56% *ee*. Slow interconversion between enantiomeric chair conformations may be responsible for the decreased enantioselectivity because using only 0.17 equiv of TCCA (0.5 oxidizing equivalents) resulted in enhanced enantioselectivity of the corresponding lactol (72% *ee*). Similarly, urea **L13**, a valuable intermediate in the chemical synthesis of biotin (*41*), was initially accessed in 60% *ee*. We observed a kinetic resolution of intermediate **H13**, which preferentially oxidized the minor enantiomer to **L13** and allowed for the isolation of **H13** in 52% yield and 97% *ee* when using 0.17 equiv TCCA. The oxidation of racemic hemiacetal **H13** by using **P7** revealed a selectivity factor of 8.3 for the oxidation to the lactone **L13**.

Catalyst **P7** exhibited high chemoselectivity for oxidation of diols over lactols (Fig. 4B), which enabled direct access to these valuable synthons by lowering the oxidant stoichiometry. We showcased this method in the concise formal synthesis of two pharmaceutical candidates, improving on previous routes in step count. Thus, we obtained tricyclic **T25**, a key intermediate in the synthesis of an HIV protease inhibitor (*42*), through oxidation of acetonide **S25** followed by one-pot acid-mediated transacetalization in 54% overall yield with 83% *ee* (Fig. 3B). In a second example, we obtained lactol **H8**, an intermediate in the synthesis of thromboxane receptor antagonist ifetroban (*43*, *44*) that is currently in clinical trials, on 2 mmol scale in 96% *ee* and 86% yield through an expedient two-step process from commercial *meso*-acid anhydride **A8**. By contrast, traditional syntheses required enantioselective methanolysis of **A8** to form monoester **M8** followed by a sequence of redundant functional group interconversions and redox manipulations (*45*, *46*).

To establish the upper limit of catalytic turnover, we surveyed decreased catalyst loadings using diethyl ketal **S17** and found that even 0.001 mol % catalyst (10 parts per million; 200,000% decrease from standard 2 mol % conditions) promoted the reaction without any loss in yield or enantioselectivity. Thus, **P7** achieved ~1.2×10^5 turnovers, which to our knowledge is the highest number recorded for an oligopeptide organocatalyst (*47*). The catalyst could also be recycled (>95% isolated from the reaction as the **P7** derived hydroxylamine) and reused (up to three times), as shown in the synthesis of **H8**, further demonstrating the potential practicality of the reported protocol (Fig. 3B and supplementary materials, section 13).

Electrochemical oxidation to enhance functional group compatibility

The optimal conditions established by using catalyst **P7** showed a broad scope with respect to the skeletal diversity of substrates, but the reaction generality was somewhat limited by compatibility of ancillary functional groups with TCCA. Substrates containing alkenes were chlorinated, resulting in diminished yields (0 to 25%). Such products are nonetheless useful synthetic intermediates; for example, unsaturated **L27** was leveraged in the preparation of another thromboxane-A_2 antagonist (*48*). To address this issue, an electrochemical protocol was developed. The key was identification of a functional group-tolerant sacrificial oxidant that would not decompose sensitive substrates but would readily participate in the cathodic half reaction (*49*). In traditional anodic oxidation reactions, protic acids or alkyl halides are used as sacrificial oxidants (*50*), but these species proved incompatible with our reaction system, likely because of the liberation of Lewis basic anions during their cathodic reduction, which led to detrimental catalyst binding. Rapid screening by using parallel reactor HTe^-Chem led to the identification of two classes of oxidants that provided high yields and enantioselectivities: dialkyl azodicarboxylates and disubstituted peroxides. We found that diisopropyl azodicarboxylate (DIAD) proved to be optimal, providing products **L27** to **29** in good yields and high enantioselectivity, with DIAD reduced to innocuous diisopropyl hydrazine-1,2-dicarboxylate (Fig. 3C). In this case, we used either an organic base 2,6-di-*tert*-butylpyridine or an inorganic buffer that consisted of $NaHCO_3$ and CO_2.

Mechanistic studies

We next conducted studies to gain understanding of the mechanistic underpinnings of the observed substrate promiscuity and high

enantioselectivity (Fig. 4A). We examined the composition of the resting state of the catalyst using in situ diffusion-ordered spectroscopy (DOSY) NMR experiments and direct analysis in real time (DART) high-resolution mass spectrometry, showing the formation of an oxoammonium-alkoxide complex (**C**) between the catalyst and the diol substrate. Further kinetic isotope effect (KIE) studies support that the rate-determining step of the reaction is a Cope-elimination from the resting state adduct to provide the desymmetrized aldehyde intermediates (**A**) (*13*). Subsequently, we performed a suite of kinetic experiments to probe additional catalyst-substrate interactions. Intermolecular competition studies between diol **S9** and mono-ol, cyclohexylmethanol, established excellent correlation [coefficient of determination (R^2) = 0.95] between the enantioenrichment of **L9** ($\Delta\Delta G^{\ddagger}$) and the relative reaction rate of **S9** versus mono-ol [$\ln(k_{\text{diol}}/k_{\text{mono-ol}})RT$] (Fig. 2B). More enantioselective catalysts showed a higher degree of selectivity toward diol over mono-ol oxidation with up to a 26-fold rate difference. These findings depict a structure in which the spectator alcohol participates in hydrogen bonding with a functional group on the peptide backbone (Fig. 4A), in addition to the covalent association of the reacting alcohol and the oxoammonium catalytic unit.

Consistent with this, product features can be used to quantitatively predict when a substrate should perform well in this reaction. We initially applied a classification algorithm to the optimization set (**S1** to **S15**) (*51*). Using a cutoff value of 85% *ee* under standard reaction conditions, substrates can be binned on the basis of the difference between the buried volumes (V_{bur}%) of each face of the lactone structure (Fig. 4C). We refined this analysis using the additional test set diols (**S16** to **S18**, **S21, S22**, and **S25** to **S29**), which were selected with the goal of assessing substrate diversity. These diols covered a broad range of ΔV_{bur}%, which revealed that substrates with a minimum ΔV_{bur}% at 3.5 Å of greater than 9.6% provide excellent enantioselectivity. We used diols **S19** and **S20** that contain a cyclobutane backbone not represented in the training data to validate the model; both **S19** and **S20** were correctly predicted to give high enantioselectivity, which suggests that this classification model can be used prospectively.

Taken together, positing that the diol motif provides multipoint binding with the catalyst, the substituents on the substrate backbone impart steric differentiation between the opposing diastereomeric transition states, providing a basis for the observed high enantioselectivity. This model of stereoinduction is simple yet highly effective, requiring only two alcohol groups that are necessary for the reaction and a relatively small steric distinction (ΔV_{bur}% = 9.6) between the two enantiotopic faces of the diol. The enantioselectivity of the reaction is otherwise insensitive to the chain length, skeletal constitution, functional groups, and other structural factors. This stereochemical model serves as the basis of the observed substrate generality.

REFERENCES AND NOTES

1. T. P. Yoon, E. N. Jacobsen, *Science* **299**, 1691–1693 (2003).
2. E. N. Jacobsen, A. Pfaltz, H. Yamamoto, *Comprehensive Asymmetric Catalysis* (Springer, 1999), vol. I to III.
3. D. A. Strassfeld, R. F. Algera, Z. K. Wickens, E. N. Jacobsen, *J. Am. Chem. Soc.* **143**, 9585–9594 (2021).
4. C. C. Wagen, S. E. McMinn, E. E. Kwan, E. N. Jacobsen, *Nature* **610**, 680–686 (2022).
5. X. Gao, H. B. Kagan, *Chirality* **10**, 120–124 (1998).
6. T. Satyanarayana, H. B. Kagan, *Adv. Synth. Catal.* **347**, 737–748 (2005).
7. H. Kim *et al.*, *Nat. Commun.* **10**, 770 (2019).
8. K. Burgess, H.-J. Lim, A. M. Porte, G. A. Sulikowski, *Angew. Chem. Int. Ed.* **35**, 220–222 (1996).
9. A. J. Metrano *et al.*, *Chem. Rev.* **120**, 11479–11615 (2020).
10. Y. Wang *et al.*, *Chem. Rev.* **121**, 12384–12444 (2021).
11. P. A. Lichtor, S. J. Miller, *Nat. Chem.* **4**, 990–995 (2012).
12. F. Formaggio *et al.*, *Chemistry* **8**, 84–93 (2002).
13. J. E. Nutting, M. Rafiee, S. S. Stahl, *Chem. Rev.* **118**, 4834–4885 (2018).
14. J. C. Siu, J. B. Parry, S. Lin, *J. Am. Chem. Soc.* **141**, 2825–2831 (2019).
15. J. M. Bobbitt, C. BrüCkner, N. Merbouh, in *Organic Reactions* (Wiley, 2010); pp. 103–424.
16. S. Nagasawa, Y. Sasano, Y. Iwabuchi, *Heterocycles* **105**, 61–114 (2022).
17. M. F. Semmelhack, C. R. Schmid, *J. Am. Chem. Soc.* **105**, 6732–6734 (1983).
18. F. Wang, M. Rafiee, S. S. Stahl, *Angew. Chem. Int. Ed.* **57**, 6686–6690 (2018).
19. Z. Ma, Q. Huang, J. M. Bobbitt, *J. Org. Chem.* **58**, 4837–4843 (1993).
20. K. Murakami *et al.*, *J. Am. Chem. Soc.* **136**, 17591–17600 (2014).
21. M. Tomizawa, M. Shibuya, Y. Iwabuchi, *Org. Lett.* **11**, 1829–1831 (2009).
22. S. D. Rychnovsky, T. L. McLernon, H. Rajapakse, *J. Org. Chem.* **61**, 1194–1195 (1996).
23. Y. Iwabuchi, *Chem. Pharm. Bull.* **61**, 1197–1213 (2013).
24. H. Shimizu, S. Onitsuka, H. Egami, T. Katsuki, *J. Am. Chem. Soc.* **127**, 5396–5413 (2005).
25. T. Suzuki, K. Morita, Y. Matsuo, K. Hiroi, *Tetrahedron Lett.* **44**, 2003–2006 (2003).
26. Y. Kashiwagi *et al.*, *Chem. Commun.* **2003**, 114–115 (2003).
27. A. Borissov *et al.*, *Chem. Soc. Rev.* **45**, 5474–5540 (2016).
28. H. Tanaka, Y. Kawakami, K. Goto, M. Kuroboshi, *Tetrahedron Lett.* **42**, 445–448 (2001).
29. D. Zell *et al.*, *J. Am. Chem. Soc.* **143**, 19078–19090 (2021).
30. S. K. Kariofillis *et al.*, *J. Am. Chem. Soc.* **144**, 1045–1055 (2022).
31. E. G. Hutchinson, J. M. Thornton, *Protein Sci.* **3**, 2207–2216 (1994).
32. A. J. Metrano *et al.*, *J. Am. Chem. Soc.* **139**, 492–516 (2017).
33. J. M. Crawford, E. A. Stone, A. J. Metrano, S. J. Miller, M. S. Sigman, *J. Am. Chem. Soc.* **140**, 868–871 (2018).
34. B. M. Trost, B. Vidal, M. Thommen, *Chemistry* **5**, 1055–1069 (1999).
35. T. Kametani, T. Suzuki, S. Kamada, K. Unno, *J. Chem. Soc., Perkin Trans. 1* **1981**, 3101–3105 (1981).
36. S. Hagishita, K. Seno, *Chem. Pharm. Bull. (Tokyo)* **37**, 327–335 (1989).
37. T. Olejniczak, F. Boratyński, A. Białońska, *J. Agric. Food Chem.* **59**, 6071–6081 (2011).
38. B. Thierry, D. Patrice, France patent FR2972453 (A1) (Minakem, France, 2012).
39. L. E. Löffler, C. Wirtz, A. Fürstner, *Angew. Chem. Int. Ed.* **60**, 5316–5322 (2021).
40. C. Zhang *et al.*, *ACS Sustain. Chem.& Eng.* **4**, 4396–4402 (2016).
41. M. Seki, Y. Takahashi, *Org. Process Res. Dev.* **25**, 1950–1959 (2021).
42. A. K. Ghosh *et al.*, *J. Med. Chem.* **63**, 4867–4879 (2020).
43. L. Rosenfeld, G. J. Grover, C. T. Stier Jr., *Cardiovasc. Drug Rev.* **19**, 97–115 (2001).
44. R. N. Misra *et al.*, *J. Med. Chem.* **36**, 1401–1417 (1993).
45. J. Das, M. F. Haslanger, J. Z. Gougoutas, M. F. Malley, *Synthesis* **1987**, 1100–1103 (1987).
46. Y. Chen, S.-K. Tian, L. Deng, *J. Am. Chem. Soc.* **122**, 9542–9543 (2000).
47. F. Giacalone, M. Gruttadauria, P. Agrigento, R. Noto, *Chem. Soc. Rev.* **41**, 2406–2447 (2012).
48. F. Lieb, U. Niewöhner, D. Wendisch, *Liebigs Ann. Chem.* **1987**, 607–615 (1987).
49. J. Rein *et al.*, *ACS Cent. Sci.* **7**, 1347–1355 (2021).
50. X. Dong, J. L. Roeckl, S. R. Waldvogel, B. Morandi, *Science* **371**, 507–514 (2021).
51. S. H. Newman-Stonebraker *et al.*, *Science* **374**, 301–308 (2021).

ACKNOWLEDGMENTS

We thank K. D. Logan, L. Alperstein, and J. S. K. Ho for substrate preparation; F. Menges (Yale University) for high-resolution mass spectrometry analysis of peptides; and H. Yoon (AbbVie), Z. Girvin (Pfizer), T. Morack (Yale University), and J. Robins (Yale University) for helpful discussion. **Funding:** This work was supported by National Institute of General Medical Sciences grant R01GM134088 (J.R., S.B.Z., J.C.S., and S.L.); National Institute of General Medical Sciences grant R35GM132092 (S.D.R., O.C.L., and S.J.M.); National Institute of General Medical Sciences grant R35GM136271 (M.A.H. and M.S.S.); ERP-Fellowship from the Studienstiftung des Deutschen Volkes (J.R.); Wiberg Graduate Research Fellowship in Chemistry (S.D.R.); National Science Foundation Graduate Research Fellowship Program (O.C.L. and S.B.Z.); and American Cancer Society–Roaring Fork Valley Research Circle Postdoctoral Fellowship PF-21-089-01-ET (M.A.H.). This work made use of the Cornell University NMR Facility, which is supported in part by the NSF through MRI award CHE-1531632. The computational portion of this work was supported by the Center for High Performance Computing (CHPC) at the University of Utah. **Author contributions:** J.R. and S.D.R. contributed equally and are listed in alphabetical order. O.C.L. and S.B.Z. contributed equally and are listed in alphabetical order. S.L., S.J.M., and M.S.S. supervised the project. J.R., S.D.R., O.C.L., S.B.Z., and J.C.S. performed synthetic experiments. J.R. and M.A.H. performed mechanistic and computational studies. B.Q.M. collected and analyzed x-ray crystallography data. The manuscript was written by S.D.R., J.R., M.A.H., M.S.S., S.J.M., and S.L. and was edited and approved by all authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Full x-ray structural data are available free of charge from the Cambridge Crystallographic Data Centre (CCDC) under deposition nos. 2207088, 2207089, 2207090, 2207091, and 2207363. All other data are available in the main text or the supplementary materials.

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adf6177
Materials and Methods
Supplementary Text
Figs. S1 to S52
Tables S1 to S43
References (*52*–*91*)
Data S1 and S2

Submitted 2 November 2022; resubmitted 16 February 2023
Accepted 29 March 2023
10.1126/science.adf6177

CHEMICAL PHYSICS

Femtosecond symmetry breaking and coherent relaxation of methane cations via x-ray spectroscopy

Enrico Ridente[1,2]†, Diptarka Hait[1,2]†‡, Eric A. Haugen[1,2], Andrew D. Ross[1,2]§, Daniel M. Neumark[1,2], Martin Head-Gordon[1,2], Stephen R. Leone[1,2,3]*

Understanding the relaxation pathways of photoexcited molecules is essential to gain atomistic-level insight into photochemistry. We performed a time-resolved study of ultrafast molecular symmetry breaking through geometric relaxation (Jahn-Teller distortion) on the methane cation. Attosecond transient absorption spectroscopy with soft x-rays at the carbon K-edge revealed that the distortion occurred within 10 ± 2 femtoseconds after few-femtosecond strong-field ionization of methane. The distortion activated coherent oscillations in the asymmetric scissoring vibrational mode of the symmetry-broken cation, which were detected in the x-ray signal. These oscillations were damped within 58 ± 13 femtoseconds because vibrational coherence was lost with the energy redistributing into lower-frequency vibrational modes. This study completely reconstructs the molecular relaxation dynamics of this prototypical example and opens avenues for exploring complex systems.

Chemical reactions arise from the motion of atomic nuclei. Atomic displacements can be described in terms of either fluctuations about a local minimum of energy or relaxation toward such a minimum from a nonequilibrium configuration. The latter often results from interaction with light because photon absorption can lead to excited electronic states with minimum energy geometries quite distinct from the initial starting point. The nonequilibrium configurations arising from light-matter interaction thus can have substantial surplus potential energy, which can drive chemical transformations. Therefore, intramolecular relaxation dynamics of photoexcited molecules are of fundamental photochemical interest.

Jahn-Teller (JT) distortion (*1*, *2*) is a special type of relaxation mechanism that spontaneously reduces the spatial symmetry of nonlinear molecules in degenerate electronic states. Molecular geometries where multiple electronic states are isoenergetic are not stable for any of the associated states (*1*) and represent a fundamental breakdown of the Born-Oppenheimer approximation (*3*). It therefore becomes energetically favorable to undergo distortions that lift the degeneracy by breaking spatial symmetry. JT distortions are ubiquitous in solids (*4*) and gas-phase molecules (*5*). In this work, we used attosecond x-ray transient absorption spectroscopy (XTAS) to study symmetry breaking of the methane cation (CH_4^+) generated through vertical strong-field ionization (SFI). This work unequivocally revealed the role and timescale of JT-induced dynamics experimentally and provided an understanding of relaxation mechanisms in molecular systems.

CH_4^+ is a classic system in which JT distortions occur (*6*–*9*). The process starts with CH_4, which is the smallest stable molecule with tetrahedral (T_d) geometry. The equilibrium C-H bond distances are 1.087 Å (*10*), and all of the H-C-H bond angles are ≈109.5° because of T_d symmetry. The ground-state molecular orbitals (MOs) are shown in Fig. 1A, and the electronic configuration of neutral CH_4 is $1a_1^2 2a_1^2 1t_2^6$ ($^1A_1$). The electronic ground-state configuration of CH_4^+ at the T_d geometry is $1a_1^2 2a_1^2 1t_2^5$, which is triply degenerate ($^2T_2$) because each of the three $1t_2$ orbitals is equally likely to be singly occupied. Therefore, CH_4^+ undergoes JT distortion away from the T_d geometry to a lower symmetry C_{2v} form (*7*, *11*–*14*). This JT distortion involves t_2 and e symmetry vibrational modes of CH_4, making CH_4^+ the simplest $T_2 \otimes (t_2 + e)$ JT problem (*15*). The resulting C_{2v} equilibrium structure was computed to have two long (1.187 Å) and two short (1.083 Å) C-H bonds (Fig. 1A), which indicates antisymmetric stretching (t_2 symmetry) relative to neutral CH_4. The angle formed by the long C-H bonds is 55.0°, and the short bonds form an angle of 125.7°, representing considerable deviations from the initial T_d geometry through bending motions (of t_2 and e symmetry in CH_4). These distortions lower the energy of the doubly occupied $3a_1$ and $1b_1$ MOs (Fig. 1A) but also destabilize the $1b_2$ singly occupied MO (SOMO). The electronic ground state of CH_4^+ therefore is $^2B_2$ (*12*).

Molecular JT-distorted forms, and CH_4^+ in particular, have been extensively studied both theoretically (*15*–*18*) and experimentally (*6*, *19*). Experimentally, time resolving the JT distortion in CH_4^+ remained an open challenge (*7*, *13*) because of the ultrafast nature of the process. JT-distorted species have been characterized in photoelectron spectroscopy experiments (*6*, *20*), but such measurements lack the temporal resolution to obtain the femtosecond timescale dynamics of symmetry breaking. Baker *et al.* (*21*) used attosecond-resolution high-harmonic emission spectroscopy to report on the onset of the JT distortions in CH_4^+ and deuterated CD_4^+ up to the first 1.6 fs. The nuclear motion in those experiments, however, cannot be reconstructed at longer times, which precludes a complete analysis of the JT relaxation process and subsequent coherent motion. Coulomb explosion experiments by Li *et al.* (*22*) probed the dynamics of CH_4^+ by recording the photofragments after interaction with two time-delayed strong-field 800-nm pulses, but temporal resolution was limited by the 25-fs pulses used as pump and probe. Furthermore, their use of a multicycle pump pulse led to several additional photoproducts from higher-energy fragmentation pathways that compete with JT distortions. It has been shown that the use of shorter, few-cycle 800-nm pulses increased the relative amount of CH_4^+ by suppressing additional product channels (*23*).

XTAS, based on attosecond- and few-femtosecond-duration soft x-ray pulses generated through high-harmonic generation (*24*), has been successfully used to study ultrafast molecular relaxation processes with high structural and temporal resolution (*25*, *26*). XTAS at the carbon K-edge is therefore an ideal platform to observe few-femtosecond timescale dynamics such as those associated with the JT distortion of CH_4^+ (*27*, *28*). The x-ray probe excites C 1s electrons to unoccupied levels, such as the SOMO, or completely unoccupied antibonding or Rydberg levels. In particular, the dipole-allowed 1s → SOMO signal in CH_4^+ is expected to be energetically well resolved from other features. Geometric changes as a result of JT distortion will strongly affect the SOMO energy, which can be traced by XTAS with few-femtosecond time resolution.

In this work, we report a joint experimental and theoretical study of the symmetry-breaking JT dynamics of CH_4^+. The cations were produced from neutral methane through abrupt, few-femtosecond SFI of CH_4 with an 800-nm, few-cycle pump pulse generated by a table-top Ti:sapphire laser. The induced dynamics were then probed with XTAS using high-harmonic–generated soft x-ray pulses at the C K-edge obtained with a 1300-nm source (*24*). The nonperturbative nature of SFI leads to an ionization window that is temporally much narrower than the 5-fs width of the pump pulse (*29*). We observed a substantial energy shift in the XTAS signal immediately upon ionization because of JT distortion. The C_{2v} minimum geometry was

[1]Department of Chemistry, University of California, Berkeley, CA 94720, USA. [2]Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [3]Department of Physics, University of California, Berkeley, CA 94720, USA.
*Corresponding author. Email: srl@berkeley.edu
†These authors contributed equally to this work.
‡Present address: Department of Chemistry and PULSE Institute, Stanford University, Stanford, CA 94305, USA.
§Present address: TOPTICA Photonics Inc., Pittsford, NY 14534, USA.

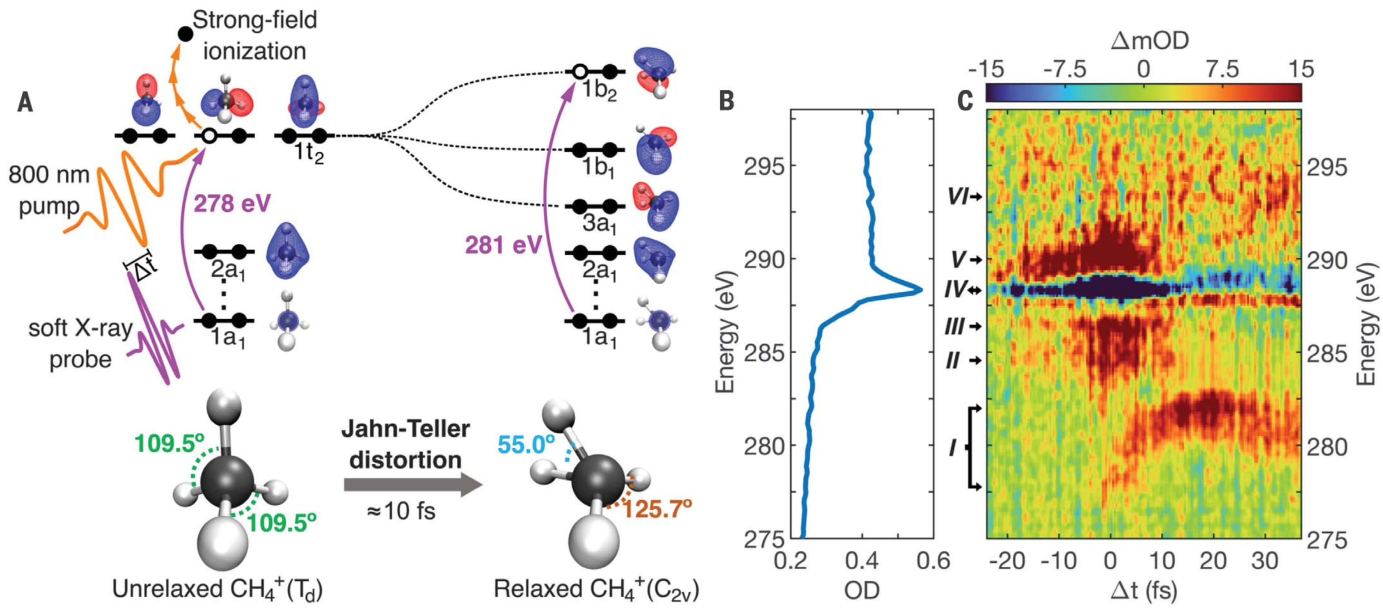


Fig. 1. XTAS measurements of CH_4^+ dynamics. (**A**) Schematic of the pump-probe process. Pump-induced SFI produces T_d symmetry CH_4^+, which undergoes JT distortion toward a C_{2v} minimum. MOs for both structures are also shown (energies not to scale). The $1a_1$ orbital is the C 1s core-level, and the remainder are $\sigma_{C\text{-}H}$ bonding orbitals. The dynamics are mapped by the 1s → SOMO probe transition induced by a time-delayed (Δt) soft x-ray probe. (**B**) Ground-state carbon K-edge x-ray absorption spectrum for CH_4. (**C**) Transient x-ray absorption at Δt < 35 fs. Negative time indicates probe preceding pump. The low-energy signal (278 to 282 eV) corresponds to the 1s → SOMO probe transition, which indicates that the C_{2v} minimum is reached by ≈10 fs.

attained within ≈10 fs, which was followed by two coherent oscillations in the signal that revealed large-amplitude scissoring motion. These coherent oscillations were damped out by ≈60 fs, indicating vibrational dephasing and eventual decoherence. The behavior of fully deuterated methane cations (CD_4^+) was also investigated to understand the effect of substituent masses on the JT dynamics.

General features of XTAS signal

CH_4 has a simple ground-state x-ray absorption spectrum (Fig. 1B). There is only one prominent peak (288 eV), which arises from the 1s → 3p Rydberg excitation (*30*). No other noteworthy pre-edge features were observed, and 1s ionization occurred at ≈290.8 eV (*31*).

Figure 1C shows the experimental carbon K-edge XTAS spectrum for CH_4^+ up to 35 fs after the pump pulse abruptly ionizes the molecule at time t = 0. The intensity of the pump beam was 3×10^{14} W/cm^2, which generated sufficient CH_4^+ without substantial photodissociation, as discussed in the supplementary materials. Previous investigations have confirmed that the major photoproduct for an 800-nm pump pulse of similar intensity and pulse duration is CH_4^+ (*32*). The negative (blue) transient signal (feature IV) corresponds to the depletion of neutral CH_4 and can be more clearly observed at higher values of the pump power (see the supplementary materials). At t = 0, other prominent features were the positive (red) signals at 278 to 282 eV (I), 284 eV (II), 287 eV (III), and 290 to 292 eV (V). The broad features II, III, and V had substantial temporal overlap with the pump pulse and could be attributed to the Stark effect of the pump pulse on core-excitation energies of CH_4 (supplementary materials). Their temporal width is longer than the pump pulse because the Stark shift in XTAS measurements is proportional to the cross-correlation and timing jitter between pump and probe. After t = 0, a positive feature at ≈287.5 eV could be observed. This feature arose because of Raman activation of the symmetric stretch vibrational mode of CH_4 by the pump pulse (supplementary materials), like the behavior that has been observed in other molecules (*33*, *34*).

Feature I was assigned to CH_4^+ on the basis of orbital-optimized density functional theory (OO-DFT) calculations (*35*) that revealed that this feature corresponded to the 1s → SOMO excitation of nonequilibrium T_d CH_4^+. OO-DFT indicated that other core-level excitations of CH_4^+ (such as 1s → σ* or the Rydberg levels) were above 288 eV in energy (supplementary materials). Feature VI in Fig. 1C corresponds to such excitations and could be observed at long times. However, feature VI did not show discernible time evolution. Conversely, feature I was well separated from all the other features and was particularly sensitive to changes in molecular geometry as the SOMO is of $\sigma_{C\text{-}H}$ character. The time evolution of this signal was the clearest reporter of the dynamics of CH_4^+, and the analysis below therefore focuses on it.

Relaxation dynamics of CH_4^+

The long-time experimental XTAS of the JT feature corresponding to the 1s → SOMO transition is given in Fig. 2A. The energetic average (henceforth abbreviated as CM1, for the first central moment) of the differential absorption [change in milli–optical density (ΔmOD)] signal (solid black line) showed three main characteristics. A rapid blue shift in energy from 278 eV (at t ≈ 0) to ≈282 eV (at t ≈ 18 fs) was followed by damped oscillations until t ≈ 60 fs and subsequently an almost time-independent signal between 281 and 281.5 eV. The width of the spectral feature increased considerably starting around t ≈ 10 fs, leading to a very broad signal at longer times.

We interpreted the behavior of this signal using quasiclassical ab initio molecular dynamics (AIMD) (*36*) trajectories on CH_4^+. Figure 2B shows the XTAS signal computed using OO-DFT from 255 AIMD trajectory geometries at different time points, revealing good agreement with the experimental results. This comparison indicates that the trajectories underlying the spectrum are a good reporter of the molecular dynamics under the experimental conditions. It is worth noting that all of the AIMD calculations were performed on the electronic ground state of CH_4^+ (i.e., were adiabatic). Nonadiabatic effects from higher-energy electronic states of CH_4^+ could potentially contribute to the small differences between experiment and theory, but the experimental evidence suggests that the system is always in

the ground electronic state under experimental conditions (supplementary materials). In general, no explicit signature of nonadiabaticity was observed in the experimental spectrum, as made evident by the good agreement between Fig. 2, A and B. The potential role of nonadiabatic effects on other aspects related to the JT distortion of CH_4 has been considered elsewhere (*9*, *15–18*).

The timescale for the JT process can be directly estimated by the time taken by the CM1 to attain the 281.5-eV value that we associate with the C_{2v} form of CH_4^+. This value was chosen because it corresponds to the asymptotic long-time limit of the experimental signal CM1, and this value was also estimated by OO-DFT to be the upper bound for the 1s → SOMO excitation of CH_4^+ at the C_{2v} equilibrium geometry (supplementary materials). This value was measured at 10 ± 2 fs from experiment by fitting an error function to the first 20 fs of the 1s → SOMO feature (supplementary materials). OO-DFT calculations find this time to be 9.4 ± 0.3 fs, confirming that JT relaxation occurs on a timescale associated with high-frequency vibrational motion.

In general, the time evolution of the signal corresponds to atomic motions associated with the relaxation process, with the oscillatory patterns suggesting involvement of vibrational modes of CH_4^+. This evolution could be analyzed further using a Fourier transform (FT) of the CM1 position from both theory and experiment (Fig. 2C). Even though the FT features were broad owing to the rapid decay in the oscillation amplitude, it was possible to identify critical frequencies. The most intense peak in the FT was at ≈1200 cm^{-1}, a frequency that corresponded to a computed normal mode of the C_{2v} minimum associated with scissoring about the H-C-H bond angles. This mode is an asymmetric scissoring mode (supplementary materials), where the scissoring motion about the smallest bond angle (i.e., the angle between the two long C-H bonds) is opposite in direction to the scissoring motion about the largest bond angle (i.e., the angle between the two short C-H bonds). However, caution must be taken in interpreting the FT features in terms of the fundamental frequencies of the C_{2v} minimum; the CH_4^+ ground-state surface has 12 distinct C_{2v} minima (*37*) and multiple seams corresponding to electronic state degeneracies, resulting in a highly anharmonic potential energy surface (PES). The FT is nonetheless an indication of several molecular motions that affect the signal and the associated timescales. The damping rate for the 1200 cm^{-1} frequency could also be estimated by fitting to the time domain experimental CM1 (Fig. 2D), revealing a lifetime of 58 ± 13 fs for the oscillations.

We used the AIMD trajectories to uncover the origins of the signal oscillations in the x-ray spectra. The trajectories indicated that the C-H

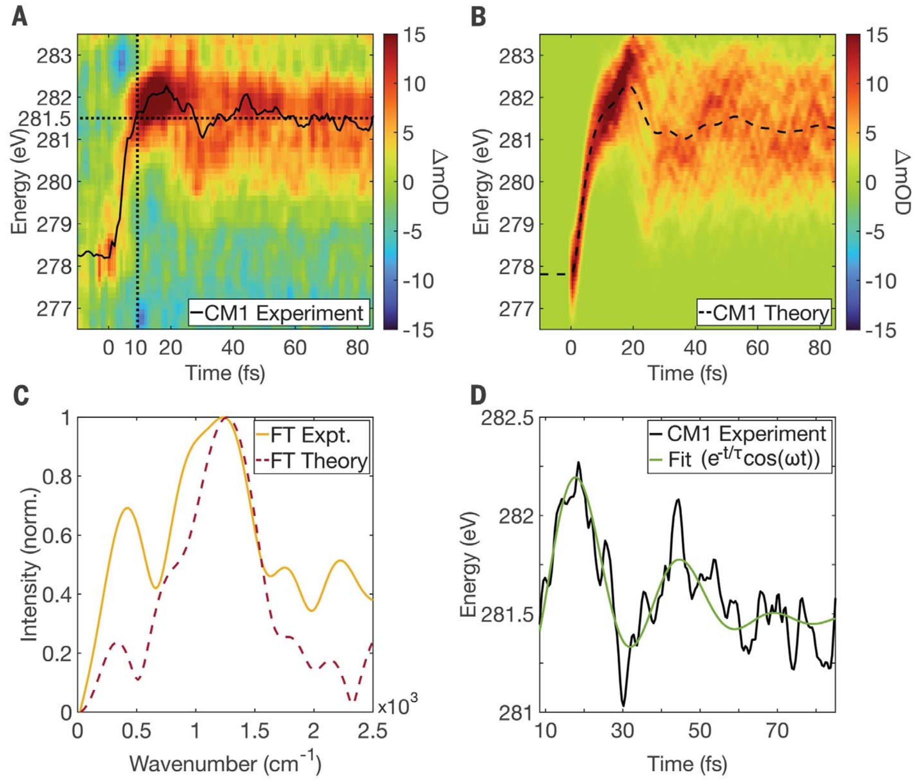


Fig. 2. Time evolution of the 1s → SOMO transition of CH_4^+. (**A**) Experimental XTAS for the 1s → SOMO transition of CH_4^+, with the first central moment (CM1) shown with the solid black line. The dotted black lines indicate the asymptotic long-time signal CM1 (281.5 eV), which corresponds to the vibrationally hot C_{2v} cation generated by the experiment, and at what time this energy is first reached (10 fs, JT timescale). (**B**) Theoretical XTAS for the same excitation. (**C**) FT of the CM1 from experiment (solid line) and theory (dashed line). The theoretical intensities have been uniformly scaled to match experiment for peak absorbance. (**D**) Experimental CM1 fit with $-e^{-t/\tau} \times \cos(\omega t)$, with $\omega = 1200$ cm^{-1}, indicating a damping lifetime of $\tau = 58 \pm 13$ fs for the vibrational dephasing.

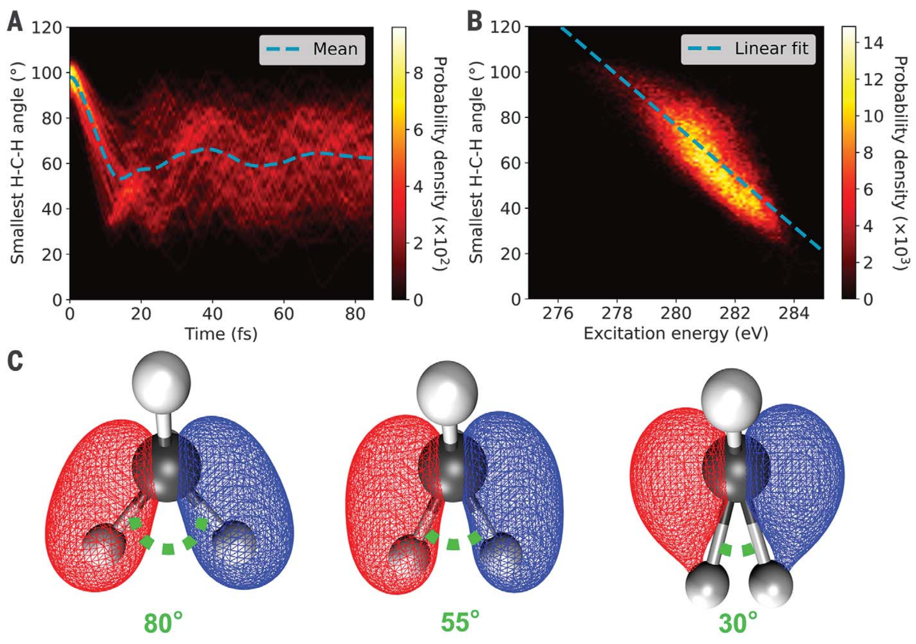


Fig. 3. Role of the smallest H-C-H angle on XTAS signal. (**A**) Time evolution of the smallest bond angle over the trajectories. (**B**) Correlation between the computed 1s → SOMO excitation energies versus the smallest bond angle of the corresponding structures. (**C**) Evolution of the SOMO with change in the smallest bond angle.

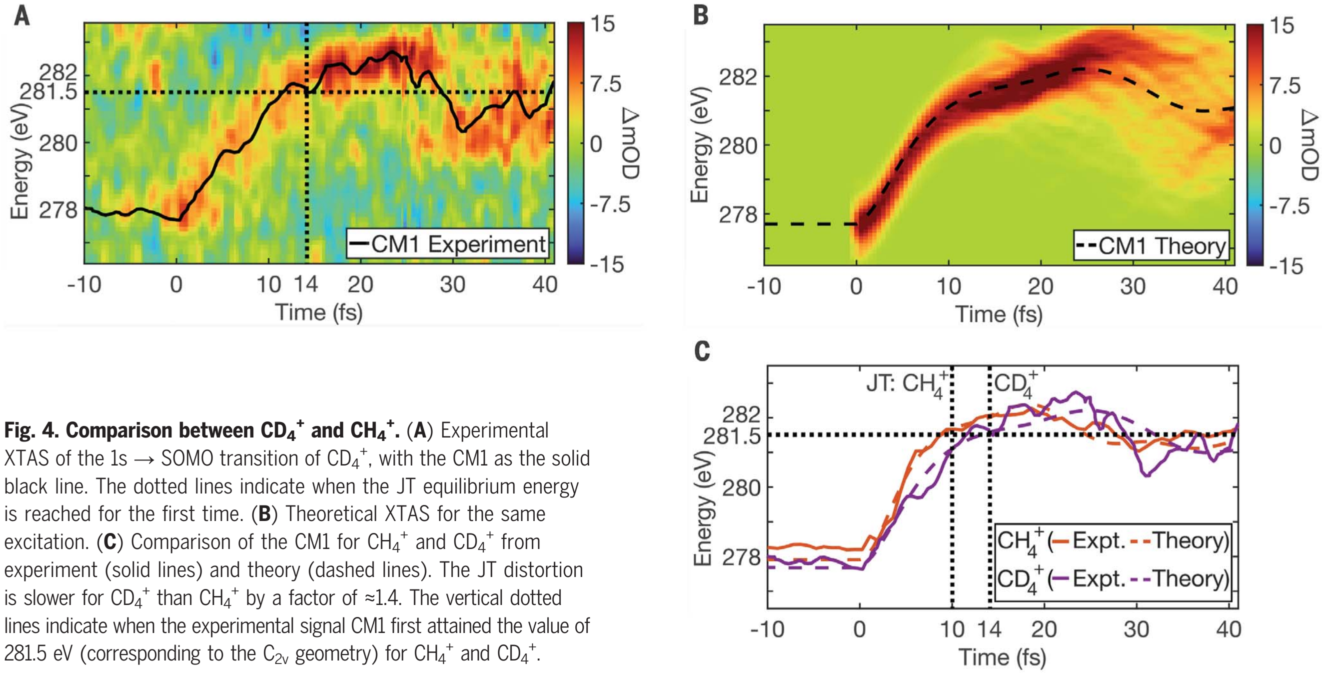


Fig. 4. Comparison between CD_4^+ and CH_4^+. (**A**) Experimental XTAS of the 1s → SOMO transition of CD_4^+, with the CM1 as the solid black line. The dotted lines indicate when the JT equilibrium energy is reached for the first time. (**B**) Theoretical XTAS for the same excitation. (**C**) Comparison of the CM1 for CH_4^+ and CD_4^+ from experiment (solid lines) and theory (dashed lines). The JT distortion is slower for CD_4^+ than CH_4^+ by a factor of ≈1.4. The vertical dotted lines indicate when the experimental signal CM1 first attained the value of 281.5 eV (corresponding to the C_{2v} geometry) for CH_4^+ and CD_4^+.

bond lengths oscillate on a timescale roughly twice as fast as the principal oscillation in the CM1 (supplementary materials). The angular oscillations were slower than the stretches, with Fig. 3A showing that the mean of the smallest molecular bond angle over all trajectories underwent a damped oscillatory motion on the same timescale as the predicted and observed XTAS signal CM1 (Fig. 2, A and B). The computed bond angle distribution also broadened rapidly over time after $t \approx 10$ fs, similar to the XTAS signal. A direct correlation between the observed signal and the smallest bond angle is revealed in Fig. 3B, which plots the smallest bond angle of the trajectory geometries used for Fig. 2B against the computed 1s → SOMO excitation energies for those geometries. It is evident that a simple linear model could capture much of the relationship between the two quantities. The excitation energy had weaker correlation with other bond angles and bond lengths (supplementary materials). Theory therefore indicated that the most important contribution to the observed time evolution of the XTAS absorption energy was from the dynamics of the smallest bond angle, to the extent that the signal could be interpreted in terms of a single parameter.

This connection can be understood using a simple orbital model (Fig. 3C). The SOMO at the C_{2v} symmetry minimum geometry of CH_4^+ is the bonding orbital arising from mixing between a C 2p orbital and a symmetry-adapted linear combination (SALC) of the 1s orbitals corresponding to H atoms in the long bonds. This SALC has antibonding character because the two H 1s orbitals have opposite phases. For small bond angles, the H 1s SALC has poorer overlap with the C 2p level as it gets closer to the nodal plane of the latter. This behavior leads to a weaker interaction and therefore lowers the mixing between the C and H centered orbitals. Furthermore, smaller angles lead to decreased H-H distance, elevating the energy of the H 1s SALC because of the local antibonding character. The resulting $\sigma_{C\text{-}H}$ bonding orbital therefore has greater nonbonding (pure C 2p) character as the angle decreases, made evident visually by the 30° angle case in Fig. 3C. Conversely, larger bond angles lead to a more stabilized SOMO with greater contribution from H orbitals. This picture is consistent with the observed increase in the 1s → SOMO x-ray probe excitation energy with decreasing bond angle shown in Fig. 3B. The x-ray oscillator strength also increased with a decrease in bond angle (supplementary materials). This behavior of the oscillator strength highlighted the increase in C 2p character of the SOMO because transitions from the C 1s level to other possible valence atomic orbitals that may contribute to the SOMO, such as C 2s or H 1s, have negligible oscillator strength. Therefore, the XTAS signal revealed the extent to which the SOMO lost C-H bonding character during the relaxation process.

It is thus apparent that the $T_d \rightarrow C_{2v}$ JT distortion activates scissoring motion about the smallest bond angle in the C_{2v} minimum, which was the most evident feature in the x-ray spectra. Furthermore, the frequencies indicated in Fig. 1C suggest that the asymmetric scissoring mode is activated to a greater extent than the symmetric scissoring mode. The greater activation of the asymmetric scissoring mode is further supported by the short time (<10 fs) evolution of the molecular bond angles (supplementary materials), which revealed that the smallest H-C-H bond angle decreases over time, whereas the largest H-C-H bond angle increases in value. For a perfectly harmonic PES, the excess energy accumulated in this asymmetric scissoring mode would remain undissipated therein, leading to undamped oscillation of the XTAS signal and geometric parameters about the minimum before radiative relaxation to the vibrational ground state. However, the PES of CH_4^+ is highly anharmonic (see discussion above), and the surplus energy spreads out to all other modes. This redistribution was observed through damping in the oscillations for both the experimental XTAS signal CM1 and the mean geometric parameters from the AIMD trajectories (Figs. 2 and 3). In addition, considerable broadening of the XTAS signal was observed after the initial 10 fs, which was mirrored by an increase in the width of the probability distributions for the geometric parameters. Figure 2D indicates that the experimental XTAS CM1 oscillations had a damping lifetime of 58 fs, and oscillations in parameters computed from AIMD trajectories were mostly damped out within 60 fs (supplementary materials). It therefore appears that a large proportion of energy was transferred out of the JT-activated asymmetric scissoring mode to other internal degrees of freedom within this timescale, constituting an ultrafast example of intramolecular vibrational energy redistribution. This process can also be described in terms of dephasing of the asymmetric scissoring mode (i.e., relaxation

toward thermal equilibrium), which was ultimately reflected in the decoherence of the XTAS signal.

Comparison between experiment and theory revealed that the XTAS signal was reporting the nuclear motion over the duration of the relaxation process. It is important to consider that SFI experiments could potentially exhibit electronic relaxation dynamics without considerable nuclear motion, involving highly excited Rydberg or cationic states (*38*, *39*). To prove that the observed XTAS signal dynamics solely arose from nuclear (rather than electronic) relaxation, we performed experiments and computations for the dynamics after SFI of CD_4. Figure 4A presents the time evolution of the resulting experimental XTAS signal, and Fig. 4B shows the computed XTAS spectrum from OO-DFT on 255 AIMD trajectories of CD_4^+. The good agreement between experiment and theory further validated the ability of the computed trajectories to successfully simulate the experimental molecular dynamics. The greater mass of the deuterium isotope should slow the scissoring motion, leading to slower time evolution of the signal for CD_4^+ relative to CH_4^+. Figure 4C confirms this to be the case by comparing the time evolution of the CM1 for both CH_4^+ and CD_4^+ from experiment and theory. The time taken by the CM1 to reach the 281.5-eV value associated with the C_{2v} cation geometry was longer for CD_4^+ than for CH_4^+, and the effect of deuteration could be gauged by the ratio (CD_4^+/CH_4^+) of these times. This ratio was measured to be 1.4 ± 0.3 from experiment and calculated to be 1.44 ± 0.06 from theory (supplementary materials). The value thus calculated was close to the ratio between the asymmetric scissoring frequencies for the smallest bond angle in CH_4^+ and CD_4^+ (computed to be 1.34). However, it is important to note that all the normal modes of C_{2v} CH_4^+ have similar frequency ratios upon isotopic substitution (supplementary materials). Nonetheless, the slower dynamics observed for CD_4^+ unambiguously reveal that the signal was reporting on nuclear dynamics. We also note that Fig. 4C shows that the relaxation dynamics after JT distortion for CD_4^+ continue to be slower up to 45 fs, from both experiment and theory. Longer time behavior for the computed CD_4^+ XTAS signal is compared with CH_4^+ in the supplementary materials (up to 85 fs), revealing slower coherent oscillations in the signal CM1 for CD_4^+. We also note that Baker *et al.* (*21*) have reported that dynamics within the first 1.6 fs of ionization were a factor of 2 to 3 slower for CD_4^+ versus CH_4^+. This behavior at very short times resulted from the decay of the autocorrelation of the nuclear wave function (*9*, *21*), which is distinct from the longer time dynamics reported here involving substantial atomic displacements, therefore representing almost nonoverlapping nuclear wave functions.

Conclusions

CH_4^+ was prepared from SFI of CH_4 and probed with XTAS near the carbon K-edge with few-femtosecond time resolution. Evolution of the excitation from the C 1s level to the valence hole revealed the dynamics of JT symmetry breaking away from the parent T_d geometry, as well as subsequent coherent motion and dissipation of released energy out of active modes. All three of these aspects of intramolecular relaxation have been successfully observed and analyzed. The combination of experiment and theory revealed that the molecule first reached the JT-distorted form within 10 ± 2 fs after ionization. This distortion involved reduction of a H-C-H bond angle from 109.5° toward 55°, which was directly reported by a blue shift in the x-ray absorption signal. The JT dynamics were found to be 1.4× slower in deuterated methane on account of the larger substituent mass, proving that the observed dynamics arose from nuclear motions. The energy released by the JT distortion drove a few coherent oscillations in the activated modes before being distributed over other molecular internal degrees of freedom, leading to damping of the oscillations within 60 fs of ionization. We note that the observed behavior for CH_4^+ is distinct from that observed in previous XTAS studies (*26*, *28*) on the dynamics of CF_4^+ and CCl_4^+ because those species are highly unstable against bond dissociation. CF_4^+ has not been experimentally detected to date (*26*); metastable CCl_4^+ has been previously observed (*28*), but signals from the intramolecular relaxation pathways were unable to be disentangled from bond breaking. The subsequent coherence and dissipation of energy from the JT activated asymmetric scissoring mode to other internal degrees of freedom was only observed in CH_4^+. This work thus opens the door to studies on how ultrafast vibrational coherence influences the redistribution of excess energy in more complex systems.

REFERENCES AND NOTES

1. H. A. Jahn, E. Teller, *Proc. R. Soc. Lond. A* **161**, 220–235 (1937).
2. I. B. Bersuker, *Chem. Rev.* **101**, 1067–1114 (2001).
3. H. C. Longuet-Higgins, in *Advances in Spectroscopy*, H. W. Thompson, Ed. (Interscience Publishers, 1961), pp. 429–472.
4. I. Persson, P. Persson, M. Sandström, A.-S. Ullström, *Dalton Trans.* **7**, 1256–1265 (2002).
5. H. Köppel, L. S. Cederbaum, W. Domcke, *J. Chem. Phys.* **89**, 2023–2040 (1988).
6. A. W. Potts, W. C. Price, *Proc. R. Soc. Lond. A* **326**, 165–179 (1972).
7. L. B. Knight Jr., J. Steadman, D. Feller, E. R. Davidson, *J. Am. Chem. Soc.* **106**, 3700–3701 (1984).
8. R. Signorell, M. Sommavilla, F. Merkt, *Chem. Phys. Lett.* **312**, 139–148 (1999).
9. C. E. M. Gonçalves, R. D. Levine, F. Remacle, *Phys. Chem. Chem. Phys.* **23**, 12051–12059 (2021).
10. E. Hirota, *J. Mol. Spectrosc.* **77**, 213–221 (1979).
11. R. Signorell, F. Merkt, *J. Chem. Phys.* **110**, 2309–2311 (1999).
12. R. Signorell, F. Merkt, *Faraday Discuss.* **115**, 205–228 (2000).
13. H. J. Wörner, R. van der Veen, F. Merkt, *Phys. Rev. Lett.* **97**, 173003 (2006).
14. H. J. Wörner, X. Qian, F. Merkt, *J. Chem. Phys.* **126**, 144305 (2007).
15. T. Mondal, A. J. C. Varandas, *J. Chem. Phys.* **135**, 174304 (2011).
16. T. Mondal, A. J. C. Varandas, *J. Chem. Phys.* **137**, 214320 (2012).
17. P. Krause, S. Matsika, *J. Chem. Phys.* **136**, 034110 (2012).
18. T. Mondal, A. J. C. Varandas, *J. Chem. Phys.* **143**, 014304 (2015).
19. J. Rabalais, T. Bergmark, L. Werme, L. Karlsson, K. Siegbahn, *Phys. Scr.* **3**, 13–18 (1971).
20. P. Baltzer *et al.*, *Chem. Phys.* **196**, 551–567 (1995).
21. S. Baker *et al.*, *Science* **312**, 424–427 (2006).
22. M. Li *et al.*, *Nat. Commun.* **12**, 4233 (2021).
23. L. Varvarezos *et al.*, *Chem. Phys. Lett.* **775**, 138687 (2021).
24. L. Barreau *et al.*, *Sci. Rep.* **10**, 5773 (2020).
25. R. Geneaux, H. J. B. Marroux, A. Guggenmos, D. M. Neumark, S. R. Leone, *Phil. Trans. R. Soc. A* **377**, 20170463 (2019).
26. Y. Pertot *et al.*, *Science* **355**, 264–267 (2017).
27. K. S. Zinchenko *et al.*, *Science* **371**, 489–494 (2021).
28. A. D. Ross *et al.*, *Chem. Sci.* **13**, 9310–9320 (2022).
29. O. Smirnova, M. Spanner, M. Ivanov, *J. Phys. B At. Mol. Opt. Phys.* **39**, S307 (2006).
30. J. Schirmer *et al.*, *Phys. Rev. A* **47**, 1136–1147 (1993).
31. J.-J. Pireaux *et al.*, *Phys. Rev. A Gen. Phys.* **14**, 2133–2145 (1976).
32. C. Wu *et al.*, *Phys. Rev. A* **75**, 043408 (2007).
33. A. Ferré *et al.*, *J. Phys. B At. Mol. Opt. Phys.* **47**, 124023 (2014).
34. K. F. Chang *et al.*, *J. Chem. Phys.* **156**, 114304 (2022).
35. D. Hait, M. Head-Gordon, *J. Phys. Chem. Lett.* **12**, 4517–4529 (2021).
36. M. Paranjothy, R. Sun, Y. Zhuang, W. L. Hase, *WIREs Comput. Mol. Sci.* **3**, 296–316 (2013).
37. M. N. Paddon-Row, D. J. Fox, J. A. Pople, K. N. Houk, D. W. Pratt, *J. Am. Chem. Soc.* **107**, 7696–7700 (1985).
38. W. G. Scherzer, H. L. Selzle, E. W. Schlag, R. D. Levine, *Phys. Rev. Lett.* **72**, 1435–1438 (1994).
39. Y. Kobayashi *et al.*, *Phys. Rev. A* **101**, 063414 (2020).
40. E. Ridente, Data for "Femtosecond Symmetry Breaking and Coherent Relaxation of Methane Cations via X-ray Spectroscopy," data set, Zenodo (2023); https://doi.org/10.5281/zenodo.7743367.

ACKNOWLEDGMENTS

Funding: This work is funded by the US Department of Energy (DOE) Office of Science, Basic Energy Sciences (BES) Program, Chemical Sciences, Geosciences, and Biosciences Division, under contract no. DE-AC02-05CH11231, through the Gas Phase Chemical Physics program (E.R., E.A.H., A.D.R., D.M.N., and S.R.L.) and the Atomic, Molecular, and Optical Sciences program (D.H. and M.H.-G.). The instrument was built with funds from the National Science Foundation through NSF MRI 1624322 and matching funds from the Lawrence Berkeley National Laboratory, the College of Chemistry, the Department of Physics, and the Vice Chancellor for Research at UC Berkeley. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the US DOE under contract no. DE-AC02-05CH11231 using NERSC award BES-ERCAP0020263. A.D.R. was additionally funded by the US DOE, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division, under contract no. DE-AC02-05-CH11231, within the Physical Chemistry of Inorganic Nanostructures Program (KC3103); by the W. M. Keck Foundation, grant no. 042982; and by the US Army Research Office under grant no. W911NF-20-1-0127. **Author contributions:** Experimental investigation: E.R., E.A.H., and A.D.R. Theoretical investigation: D.H. Experiment supervision: S.R.L. and D.M.N. Theory supervision: M.H.-G. Writing – original draft: E.R., D.H., and E.A.H. Writing – review & editing: E.R., D.H., E.A.H., A.D.R., M.H.-G., D.M.N., and S.R.L. **Competing interests:** The electronic structure calculations were performed in Q-Chem, which is partially owned by M.H.-G. The other authors declare no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials. The experimental and theoretical data can be accessed on Zenodo (*40*).

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adg4421
Materials and Methods
Figs. S1 to S25
Tables S1 to S3
References (*41*–*64*)

Submitted 23 December 2022; accepted 19 April 2023
10.1126/science.adg4421

QUANTUM OPTICS

Entangling microwaves with light

R. Sahu[1]*†, L. Qiu[1]*†, W. Hease[1]‡, G. Arnold[1], Y. Minoguchi[2], P. Rabl[2,3,4,5], J. M. Fink[1]*

Quantum entanglement is a key resource in currently developed quantum technologies. Sharing this fragile property between superconducting microwave circuits and optical or atomic systems would enable new functionalities, but this has been hindered by an energy scale mismatch of >10^4 and the resulting mutually imposed loss and noise. In this work, we created and verified entanglement between microwave and optical fields in a millikelvin environment. Using an optically pulsed superconducting electro-optical device, we show entanglement between propagating microwave and optical fields in the continuous variable domain. This achievement not only paves the way for entanglement between superconducting circuits and telecom wavelength light, but also has wide-ranging implications for hybrid quantum networks in the context of modularization, scaling, sensing, and cross-platform verification.

The ability to manipulate and measure quantum mechanical properties such as quantum superpositions and entanglement in a variety of physical systems serves as the basis for the development of quantum technologies, for which the demonstration of quantum advantage with tens of superconducting qubits (*1*), an ultracoherent quantum memory with nuclear spins (*2*), and distributed quantum entanglement over tens of kilometers using optical photons (*3*) represents just a few of the highlights that have already been achieved. Combining these techniques (*4*–*6*) will enable the realization of general-purpose quantum networks in which remote quantum nodes capable of storing and processing quantum information seamlessly communicate with each other by distributing entanglement over optical channels (*7*). Aspects of this approach have already been adopted to connect and entangle various quantum platforms remotely. These involve single atoms, ions, atomic ensembles, quantum dots, rare-earth ions, and nitrogen-vacancy centers (*8*). However, such long-distance quantum connectivity is considerably more difficult to achieve with microwave-based platforms such as semiconductor spin qubits (*9*) or local cryogenic networks of superconducting circuits (*10*, *11*), for which no natural interface to propagating optical photons is available.

To overcome this limitation, efforts are currently focused on the development of coherent quantum transducers between microwave and optical photons (*12*–*17*). Direct noiseless conversion of a quantum state typically relies on a beam-splitter process in which a strong driving field mediates the conversion between weak microwave and optical signals. However, a deterministic channel based on direct transduction has exceptionally stringent requirements on conversion efficiency and added classical noise that are still out of reach. Alternatively, quantum states can be teleported over long distances (*18*–*20*) by first establishing an entangled state between microwave and optical photons. Because they are assisted by an error-free classical signal (*21*, *22*), such teleportation-based communication channels are more tolerant to noise and losses, and their channel capacity is never zero as long as a finite amount of entanglement can be shared between the communicating nodes (*22*).

Theoretical outline and experimental setup

We generated and verified entanglement between microwave and optical fields in the continuous-variable domain (*23*) using an ultralow-noise cavity electro-optical device (*21*). Our device consists of a 5-mm-diameter, 150-μm-thick lithium niobate optical resonator placed inside a superconducting aluminum microwave cavity at a temperature of 7 mK, described in detail in (*24*). The microwave mode $\hat{a}_e$ is electro-optically coupled to the colocalized optical whispering gallery modes at $\omega_o/2\pi \approx 193.46$ THz through the Pockels effect (Fig. 1A). The microwave resonance frequency $\omega_e/2\pi$ is tuned to the optical free spectral range (FSR) of 8.799 GHz to realize a triply resonant system with interaction Hamiltonian:

$$\hat{H}_{\text{int}} = \hbar g_0 \hat{a}_p \hat{a}_e^\dagger \hat{a}_o^\dagger + h.c. \tag{1}$$

where g_0 is the vacuum electro-optical coupling rate and $\hat{a}_p$ ($\hat{a}_o$) is the annihilation operator of the optical pump (Stokes) mode (*25*–*28*). Herein, we have ignored the interaction with the suppressed optical anti-Stokes mode $\hat{a}_t$ (Fig. 1B) (*29*).

In this sideband suppressed situation, two-mode squeezing interaction was achieved with a strong resonant optical pump tone, yielding the simple effective Hamiltonian, $\hat{H}_{\text{eff}} = \hbar g_0 \sqrt{\bar{n}_p}\left(\hat{a}_e^\dagger \hat{a}_o^\dagger + \hat{a}_e \hat{a}_o\right)$, where $\bar{n}_p = \langle \hat{a}_p^\dagger \hat{a}_p \rangle$ is the mean intracavity photon number of the optical pump mode. Entanglement between the out-propagating microwave and the optical field was generated through spontaneous parametric down-conversion (SPDC) below the parametric instability threshold ($C < 1$). Here, $C = 4\bar{n}_p g_0^2/(\kappa_e \kappa_o)$ is the cooperativity with the vacuum coupling rate $g_0/2\pi \approx 37$ Hz and the total loss rates of the microwave and optical Stokes modes $\kappa_e/2\pi \approx 11$ MHz and $\kappa_o/2\pi \approx 28$ MHz. The required ultralow-noise operation for entanglement generation was achieved in the pulsed regime because of the slow heating of this millimeter-sized device despite the required high-power optical pump (*24*).

Establishing nonclassical correlations

We characterized the microwave and optical output fields using continuous variables, i.e.,

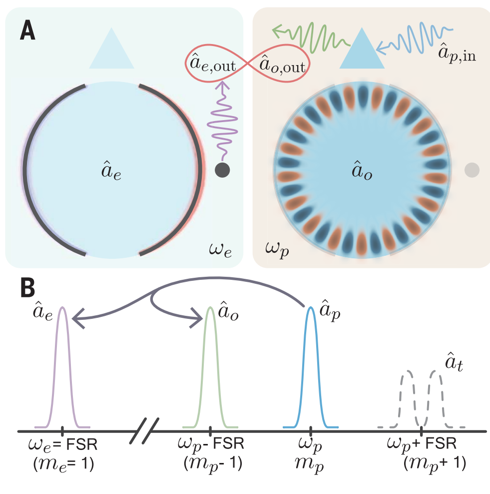


Fig. 1. Physical and conceptual mode configuration. (**A**) Simulated microwave (left) and optical (right) mode distribution with azimuthal number $m_e = 1$ and $m_o = 17$ (for illustration, experimentally $m_o \approx 20{,}000$). Phase matching is fulfilled from the condition $m_o = m_p - m_e$, and entanglement is generated and verified between the out-propagating microwave field $\hat{a}_{e,\text{out}}$ and the optical Stokes field $\hat{a}_{o,\text{out}}$. (**B**) Sketch of the density of states of the relevant modes. Under the condition $\omega_p - \omega_o = \omega_e$, the strong pump tone in $\hat{a}_p$ produces entangled pairs of microwave and optical photons in $\hat{a}_e$ and $\hat{a}_o$ through spontaneous parametric down-conversion. Frequency upconversion is suppressed through hybridization of the anti-Stokes mode $\hat{a}_t$ with an auxiliary mode.

[1]Institute of Science and Technology Austria, am Campus 1, 3400 Klosterneuburg, Austria. [2]Vienna Center for Quantum Science and Technology, Atominstitut, TU Wien, 1040 Vienna, Austria. [3]Walther-Meißner-Institut, Bayerische Akademie der Wissenschaften, 85748 Garching, Germany. [4]Technische Universität München, TUM School of Natural Sciences, 85748 Garching, Germany. [5]Munich Center for Quantum Science and Technology (MCQST), 80799 Munich, Germany.
***Corresponding author. Email: rsahu@ist.ac.at (R.S.); liu.qiu@ist.ac.at (L.Q.); jfink@ist.ac.at (J.M.F.)**
†These authors contributed equally to this work.
‡Present address: Quandela SAS, 91120 Palaiseau, France.

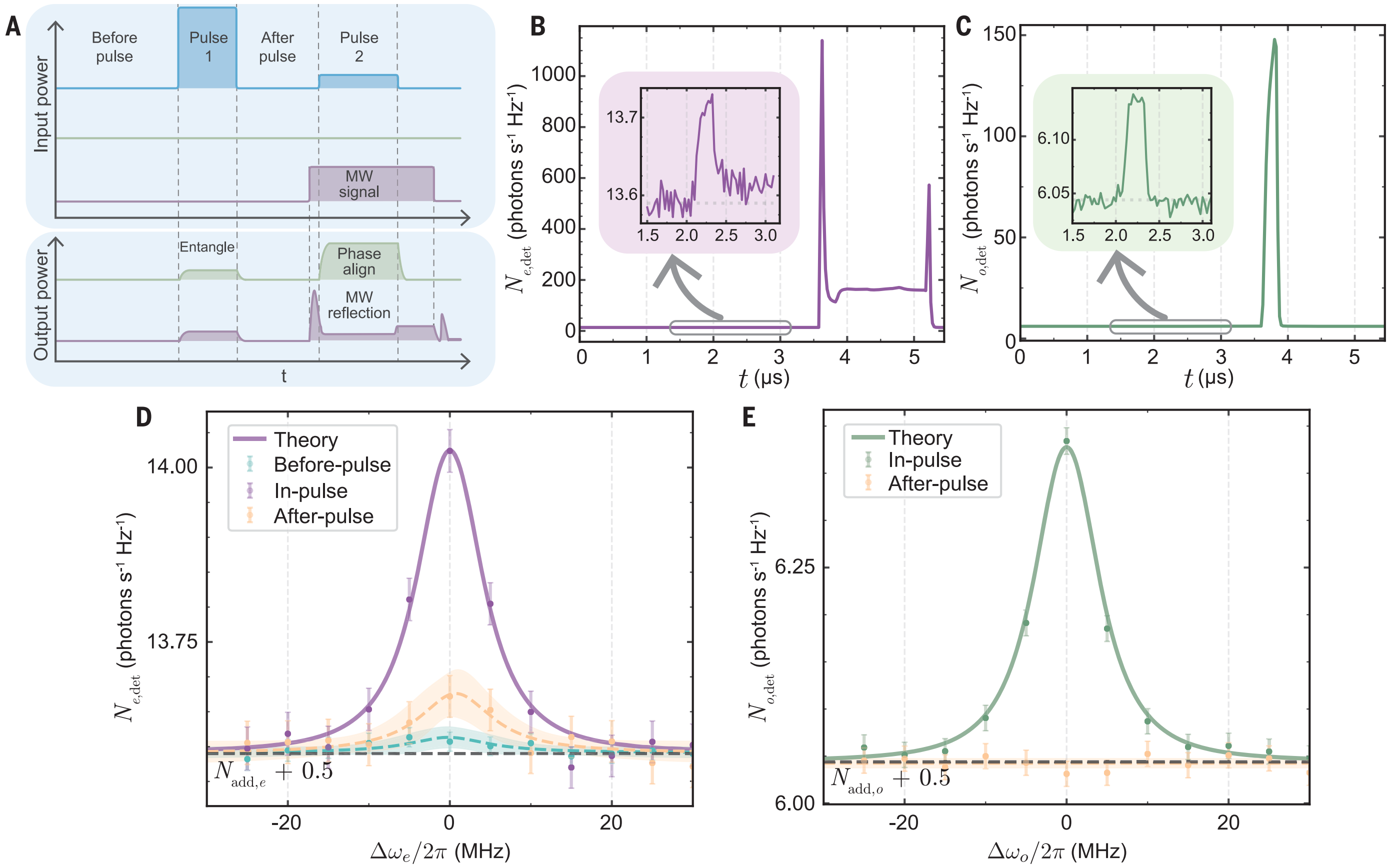


Fig. 2. Measurement sequence and noise powers. (**A**) Schematic pulse sequence of a single measurement. The optical pulse 1 is applied at ω_p and amplifies the vacuum (and any thermal noise) in the two modes $\hat{a}_e$ and $\hat{a}_o$, thus generating the SPDC signals. One micro-second later, a second optical pump with ~10 times lower power is applied, together with a coherent microwave pulse at ω_e. The microwave photons stimulate the optical pump to down-convert, which generates a coherent pulse in the $\hat{a}_o$ mode that is used to extract slow local oscillator phase drifts. (**B** and **C**) Measured output power from the $\hat{a}_e$ and $\hat{a}_o$ modes in units of photons per second in a 1-Hz bandwidth and averaged over a million experiments. The SPDC signals are shown in the insets, with the dashed gray lines indicating the calibrated detection noise floor $N_{j,\mathrm{add}}$ + 0.5. (**D**) Corresponding microwave output power spectral density versus $\Delta\omega_e = \omega - \omega_e$ centered on resonance right before the entanglement pulse, during the pulse, and right after the pulse, as indicated in (A). Yellow and green dashed lines are fits to a Lorentzian function that yields the microwave bath occupancies before and after the entangling pulse. Error bars represent the 2σ statistical SE, and the shaded regions represent the 95% confidence interval of the fit. (**E**) Corresponding optical output power spectral density versus $\Delta\omega_o = \omega_o - \omega$ during and after the entanglement pulse, both normalized to the measured noise floor before the pulse. The in-pulse noise spectra in (D) and (E) are fit jointly with theory, which yields C = 0.18 ± 0.01 and $\bar{n}_{e,\mathrm{int}} = 0.07 \pm 0.03$.

the dimensionless quadrature pairs X_j and P_j (where $j = e,o$ for microwave and optics), which satisfy the canonical commutation relations $[X_j, P_j] = i$. A pair of Einstein-Podolsky-Rosen–type operators, $X_+ = \frac{1}{\sqrt{2}}(X_e + X_o)$ and $P_- = \frac{1}{\sqrt{2}}(P_e - P_o)$, were then constructed (*30*). The microwave and optical output fields are entangled if the variance of the joint operators is reduced below the vacuum level, i.e., $\Delta^-_{\mathrm{EPR}} = \langle X_+^2 + P_-^2 \rangle < 1$. This is commonly referred to as the Duan-Simon criterion (*31*, *32*). Here, the entanglement was established deterministically between the quadratures of two propagating bosonic modes and verified through the measured quadrature variances based on all collected data (*33*). This is in contrast to probabilistic entanglement in the single photon basis, which is established with finite probability only after successful heralding from an auxiliary measurement (*34*).

For entanglement generation, we used a 250-ns-long optical pump pulse (≈154 mW, $C \approx 0.18$, $\bar{n}_p \approx 1.0 \times 10^{10}$) at a 2-Hz repetition rate (pulse 1 in Fig. 2A). The output optical signal was filtered through a Fabry-Perot cavity to reject the strong pump. The microwave output field was amplified by a high-electron-mobility transistor amplifier. Both outputs were down-converted to an intermediate frequency of 40 MHz with two local oscillators, and the four quadratures were extracted from heterodyne detection. Long-term phase stability between the two local oscillators was achieved by extraction of the relative phase drift by means of a second phase-alignment pump pulse that was applied 1 μs after each entanglement pulse, together with a coherent resonant microwave pulse (Fig. 2A). This generated a high signal-to-noise coherent optical signal through stimulated parametric down-conversion and allowed for phase alignment in each individual measurement (*29*).

Figure 2, B and C, shows the time-domain power over 1 million averages for the on-resonance microwave and optics signal with a 40-MHz measurement bandwidth. The two insets show the microwave (optical) signal from spontaneous parametric down-conversion (SPDC) from pulse 1 (compare Fig. 2A) with an emission bandwidth of ≈10 MHz. The larger signals during the second pump pulse are the reflected microwave pulse and the generated optical tone, and these were used for phase alignment. The off-resonance raw power measurements were rescaled to the

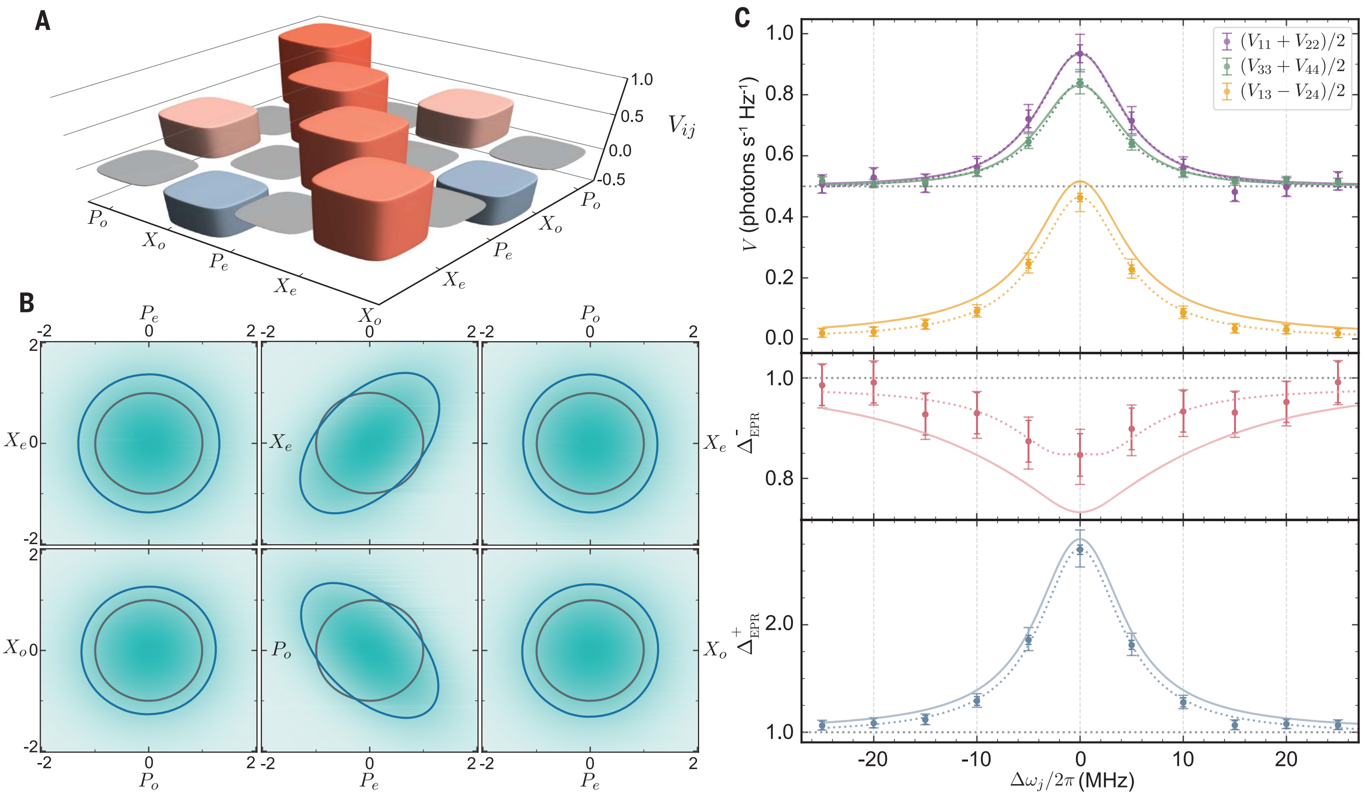


Fig. 3. Characterization of the two-mode squeezed state. (**A**) Measured covariance matrix V_{ij} in its standard form plotted for $\Delta\omega_j$ = 0 based on 925,000 measurements. (**B**) Corresponding Wigner function marginals of different output quadrature pairs compared with vacuum. The contours in blue (gray) represent the 1/*e* fall-off from the maximum for the measured state (vacuum). The middle two plots show two-mode squeezing below the vacuum level in the diagonal and off-diagonal directions. (**C**) Top, measured average microwave output noise $\bar{V}_{11} = (V_{11} + V_{22})/2$ (purple), average optical output noise $\bar{V}_{33} = (V_{33} + V_{44})/2$ (green), and average correlations $\bar{V}_{13} = (V_{13} - V_{24})/2$ (yellow) as a function of measurement detunings. Solid lines represent the joint theory fit, and dashed lines are individual Lorentzian fits to serve as a guide to the eye. Middle bottom, two-mode squeezing in red (antisqueezing in blue) calculated from the top panels as $\Delta^{\pm}_{\mathrm{EPR}} = \bar{V}_{11} + \bar{V}_{33} \pm 2\bar{V}_{13}$. The darker-colored error bars represent the 2σ statistical error, and the outer (faint) 2σ error bars also include the systematic error in calibrating the added noise of the measurement setup.

detection noise floor $N_{j,\mathrm{add}}$ + 0.5, with the added noise $N_{e,\mathrm{add}}$ = 13.1 ± 0.4 (2σ errors throughout) due to loss and amplifier noise and $N_{o,\mathrm{add}}$ = 5.5 ± 0.2 due to optical losses, which were carefully determined using noise thermometry of a temperature-controlled 50 Ω load and four-port calibration, respectively (*29*). Because the emission bandwidth was smaller than the measurement bandwidth, the time domain measurement does not reveal the full SPDC amplitude.

We continued the analysis in the frequency domain by calculating the Fourier transform of each measurement for three separate time intervals: before (2 μs), during (200 ns), and immediately after (500 ns) the entangling pump pulse. Figure 2D shows the resulting average microwave noise spectra for all three time intervals with corresponding fit curves (dashed lines) and theory (solid line). The spectra were in situ calibrated using the before-pulse off-resonance (waveguide noise) noise floor (*29*). Using independent measurements, we determined this noise floor $\bar{n}_{e,\mathrm{wg}} = 0.001 \pm 0.002$ at the low average pump power of ≈0.12 μW used in this experiment (*29*). Before and after the pump pulse, the intrinsic microwave bath occupancies were fitted to be $\bar{n}_{e,\mathrm{int}} = 0.03 \pm 0.01$ and 0.09 ± 0.03, respectively, above the vacuum level. Similarly, Fig. 2E shows the obtained average optical noise spectra during and after the pump pulse, calibrated through the shot noise level in the heterodyne measurement before the pulse. As expected, the optical noise level after the pulse returned back to the shot noise level.

During the pump pulse, approximately Lorentzian-shaped microwave and optical power spectra were generated through the SPDC process. A joint fit of the microwave and optical power spectral density during the pulse was performed using a five-mode theoretical model that includes the effects of measurement bandwidth. In this model, the in-pulse microwave bath occupancy $\bar{n}_{e,\mathrm{int}} = 0.07 \pm 0.03$ and the cooperativity *C* = 0.18 ± 0.01 are the only free fit parameters (*29*). The narrowed microwave linewidth $\kappa_{e,\mathrm{eff}}/2\pi$ = 9.8 ± 1.8 MHz (taken from a Lorentzian fit) agrees with coherent electro-optical back-action (*28*). From the extracted numbers, we conclude that during the entanglement pulse, the quantum noise dominates the intrinsic microwave thermal noise, a prerequisite for microwave-optics entanglement generation.

The quadratures X_j and P_j were extracted as a function of frequency, i.e., at $\Delta\omega_j = \pm(\omega - \omega_j)$ around the resonances due to energy conservation in the SPDC process (*29*, *35*) during the pump pulse. At each frequency, the bipartite Gaussian state of the propagating output fields was fully characterized by the 4 × 4 covariance matrix (CM) $V_{ij} = \langle \delta u_i \delta u_j + \delta u_j \delta u_i \rangle/2$, where $\delta u_i = u_i - \langle u_i \rangle$ and $u \in \{X_e, P_e, X_o, P_o\}$ (*29*). The CM corresponds to the quantum state of the propagating fields in the coaxial line and the coupling prism at the device output, i.e., before setup losses or amplification incur. The diagonal elements in *V* correspond to the individual output field quadrature variances in dimensionless units, which can be obtained by subtracting the added detection

noise offsets from the measured variances, i.e., $V_{ii}(\Delta\omega) = V_{ii,\text{meas}}(\Delta\omega) - N_{i,\text{add}}$.

The obtained CM from the data in Fig. 2 at $\Delta\omega = 0$ is shown in Fig. 3A in its standard form (*31*, *32*). The nonzero off-diagonal elements indicate strong correlations between microwave and optical quadratures. To verify the quantum correlation, the two-mode squeezed quadratures are more intuitively visualized in terms of the quasiprobability Wigner function:

$$W(\boldsymbol{u}) = \frac{\exp\left[-\frac{1}{2}\boldsymbol{u}V^{-1}\boldsymbol{u}^T\right]}{\pi^2\sqrt{\det(\mathrm{V})}} \quad (2)$$

where $\boldsymbol{u} = (X_e, P_e, X_o, P_o)$. Different marginals of this Wigner function are shown in Fig. 3B. The marginals from the same quadratures (X_e, P_e) and (X_o, P_o) show uncorrelated thermal noise above the vacuum noise (gray circle) from SPDC. The cross-quadrature marginals (X_e, X_o) and (P_e, P_o) show two-mode squeezing in the diagonal and off-diagonal directions below the vacuum level. The amount of squeezing was slightly different between the two because of the statistical uncertainty in the measured CM.

Figure 3C shows the amount of two-mode squeezing between microwave and optical output fields. The averaged microwave quadrature variance (purple dots) $\bar{V}_{11} = (V_{11} + V_{22})/2$ and the averaged optics quadrature variance (green dots) $\bar{V}_{33} = (V_{33} + V_{44})/2$ are shown in the top panel, along with the prediction from the five-mode theory (solid line) and a simple fit to a Lorentzian function (dashed line), which show perfect agreement. Measured microwave-optical correlations (yellow dots) $\bar{V}_{13} = (V_{13} - V_{24})/2$ and the Lorentzian fit (dashed line) lie slightly below the theoretical prediction (solid line), which we attribute to remaining imperfections in the phase stability (*29*).

The bottom two panels of Fig. 3C show the squeezed and antisqueezed joint quadrature variances $\Delta^{\mp}_{\text{EPR}} = \bar{V}_{11} + \bar{V}_{33} \mp 2\bar{V}_{13}$ (red and blue, respectively). We observed two-mode squeezing below the vacuum level, i.e., $\Delta^{-}_{\text{EPR}} < 1$, with a bandwidth close to the effective microwave linewidth. The maximal two-mode squeezing of $0.72^{+0.31}_{-0.25}$ dB is observed on resonance where $\Delta^{-}_{\text{EPR}} = 0.85^{+0.05}_{-0.06}$ (2σ, 95% confidence) obtained from ~1 million pulses with $\bar{V}_{11} = 0.93$, $\bar{V}_{33} = 0.84$, and $\bar{V}_{13} = 0.46$. The reported 2σ error on Δ^{-}_{EPR} takes both statistical and systematic errors into account. Thus, the value of Δ^{-}_{EPR} beats the classical limit ($\Delta^{-}_{\text{EPR}} = 1$) by >5σ (*29*). The measured two-mode squeezing signifies an itinerant microwave-optical entangled state with a logarithmic negativity of $E_N = 0.17$. The supplementary materials contain additional data for longer pulses and varying optical pump power, which corroborate the presented results and findings (*29*).

Conclusions

The demonstration of deterministic quantum entanglement between propagating microwave and optical photons establishes a nonclassical communication channel between circuit quantum electrodynamics and quantum photonics. The achieved entanglement generation rate of ≈0.11 ebits/200-ns-long pulse (*29*) is in practice limited by the slow pulse repetition rate. We expect orders-of-magnitude higher rates with improved thermalization, higher microwave and optical quality factors, and electro-optic coupling enhancements that reduce the required pump power and the associated thermal load. Coupling efficiency improvements will allow for higher levels of two-mode squeezing (*29*) and facilitate deterministic entanglement distribution schemes to qubits (*29*, *36*), teleportation-based state transfer (*21*, *22*, *29*), and quantum-enhanced remote detection (*37*). This device and state preparation scheme can also be used directly for probabilistic heralding assisted protocols (*7*, *38*, *39*) when the cooperativity is somewhat reduced. This is the most promising way forward to mitigating optical setup losses and extending the entanglement to room-temperature fiber optics. Being fully compatible with superconducting qubits in a millikelvin environment, such a device will facilitate the integration of remote superconducting quantum processors into a single coherent optical quantum network. This is relevant not only for modularization and scaling (*40*, *41*) but also for efficient cross-platform verification of classically intractable quantum processor results (*42*).

REFERENCES AND NOTES

1. F. Arute *et al*., *Nature* **574**, 505–510 (2019).
2. M. Zhong *et al*., *Nature* **517**, 177–180 (2015).
3. Y. Yu *et al*., *Nature* **578**, 240–245 (2020).
4. Z.-L. Xiang, S. Ashhab, J. Q. You, F. Nori, *Rev. Mod. Phys.* **85**, 623–653 (2013).
5. G. Kurizki *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **112**, 3866–3873 (2015).
6. A. A. Clerk, K. W. Lehnert, P. Bertet, J. R. Petta, Y. Nakamura, *Nat. Phys.* **16**, 257–267 (2020).
7. L.-M. Duan, M. D. Lukin, J. I. Cirac, P. Zoller, *Nature* **414**, 413–418 (2001).
8. S.-H. Wei *et al*., *Laser Photonics Rev.* **16**, 2100219 (2022).
9. G. Burkard, T. D. Ladd, J. M. Nichol, A. Pan, J. R. Petta, arXiv:2112.08863 [cond-mat.mes-hall] (2021).
10. P. Magnard *et al*., *Phys. Rev. Lett.* **125**, 260502 (2020).
11. Y. Zhong *et al*., *Nature* **590**, 571–575 (2021).
12. X. Han, W. Fu, C.-L. Zou, L. Jiang, H. X. Tang, *Optica* **8**, 1050 (2021).
13. M. Mirhosseini, A. Sipahigil, M. Kalaee, O. Painter, *Nature* **588**, 599–603 (2020).
14. R. Sahu *et al*., *Nat. Commun.* **13**, 1276 (2022).
15. B. M. Brubaker *et al*., *Phys. Rev. X* **12**, 021062 (2022).
16. A. Kumar *et al*., *Nature* **615**, 614–619 (2023).
17. M. J. Weaver *et al*., arXiv:2210.15702 [quant-ph] (2022).
18. S. Takeda, T. Mizuta, M. Fuwa, P. van Loock, A. Furusawa, *Nature* **500**, 315–318 (2013).
19. K. G. Fedorov *et al*., *Sci. Adv.* **7**, eabk0891 (2021).
20. S. L. N. Hermans *et al*., *Nature* **605**, 663–668 (2022).
21. A. Rueda, W. Hease, S. Barzanjeh, J. M. Fink, *npj Quantum Inf.* **5**, 108 (2019).
22. J. Wu, C. Cui, L. Fan, Q. Zhuang, *Phys. Rev. Appl.* **16**, 064044 (2021).
23. U. L. Andersen, T. Gehring, C. Marquardt, G. Leuchs, *Phys. Scr.* **91**, 053001 (2016).
24. W. Hease *et al*., *PRX Quantum* **1**, 020315 (2020).
25. M. Tsang, *Phys. Rev. A* **81**, 063837 (2010).
26. V. S. Ilchenko, A. A. Savchenkov, A. B. Matsko, L. Maleki, *J. Opt. Soc. Am. B* **20**, 333 (2003).
27. L. Fan *et al*., *Sci. Adv.* **4**, eaar4994 (2018).
28. L. Qiu, R. Sahu, W. Hease, G. Arnold, J. M. Fink, arXiv:2210.12443 [quant-ph] (2022).
29. See the supplementary materials.
30. A. Einstein, B. Podolsky, N. Rosen, *Phys. Rev.* **47**, 777–780 (1935).
31. L.-M. Duan, G. Giedke, J. I. Cirac, P. Zoller, *Phys. Rev. Lett.* **84**, 2722–2725 (2000).
32. R. Simon, *Phys. Rev. Lett.* **84**, 2726–2729 (2000).
33. Z. Y. Ou, S. F. Pereira, H. J. Kimble, K. C. Peng, *Phys. Rev. Lett.* **68**, 3663–3666 (1992).
34. H. Bernien *et al*., *Nature* **497**, 86–90 (2013).
35. S. Zippilli, G. D. Giuseppe, D. Vitali, *New J. Phys.* **17**, 043025 (2015).
36. J. Agustí, Y. Minoguchi, J. M. Fink, P. Rabl, *Phys. Rev. A* **105**, 062454 (2022).
37. S. Barzanjeh *et al*., *Phys. Rev. Lett.* **114**, 080503 (2015).
38. C. Zhong *et al*., *Phys. Rev. Lett.* **124**, 010511 (2020).
39. S. Krastanov *et al*., *Phys. Rev. Lett.* **127**, 040503 (2021).
40. S. Bravyi, O. Dial, J. M. Gambetta, D. Gil, Z. Nazario, *J. Appl. Phys.* **132**, 160902 (2022).
41. J. Ang *et al*., arXiv:2212.06167 [quant-ph] (2022).
42. J. Knörzer, D. Malz, J. I. Cirac, arXiv:2212.07789 [quant-ph] (2022).
43. R. Sahu, Data for: Entangling microwaves with light, Zenodo; https://doi.org/10.5281/zenodo.7789418.

ACKNOWLEDGMENTS

L.Q. thanks J. Li and D. Vitali for fruitful discussions. **Funding:** This work was supported by the European Research Council (grant no. 758053, ERC StG QUNNECT) and the European Union's Horizon 2020 Research and Innovation Program (grant no. 899354, FETopen SuperQuLAN). L.Q. acknowledges generous support from the ISTFELLOW program. W.H. is the recipient of an ISTplus postdoctoral fellowship with funding from the European Union's Horizon 2020 Research and Innovation Program (Marie Sklodowska-Curie grant no. 754411). G.A. is the recipient of a DOC fellowship of the Austrian Academy of Sciences at IST Austria. J.M.F. acknowledges support from the Austrian Science Fund (FWF) through BeyondC (grant no. F7105) and the European Union's Horizon 2020 Research and Innovation Program (grant no. 862644, FETopen QUARTET). **Author contributions:** R.S., W.H., L.Q., and G.A. worked on the experimental setup. R.S. and L.Q. performed the measurements. L.Q. and R.S. analyzed the data. L.Q. developed the theory with contributions from R.S., Y.M., and P.R. R.S. and L.Q. wrote the manuscript with contributions from all authors. J.M.F. supervised the project. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials or have been deposited on Zenodo (*43*).

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adg3812
Supplementary Text
Figs. S1 to S15
Tables S1 and S2
References (*44*–*54*)

Submitted 20 December 2022; accepted 19 April 2023
10.1126/science.adg3812

CALORICS

High-performance multimode elastocaloric cooling system

Suxin Qian[1,2]†, David Catalini[2]†‡, Jan Muehlbauer[2], Boyang Liu[3], Het Mevada[2], Huilong Hou[4,5,6], Yunho Hwang[2], Reinhard Radermacher[2], Ichiro Takeuchi[3,7]*

Developing zero–global warming potential refrigerants has emerged as one area that helps address global climate change concerns. Various high-efficiency caloric cooling techniques meet this goal, but scaling them up to technologically meaningful performance remains challenging. We have developed an elastocaloric cooling system with a maximum cooling power of 260 watts and a maximum temperature span of 22.5 kelvin. These values are among the highest reported for any caloric cooling system. Its key feature is the compression of fatigue-resistant elastocaloric nitinol (NiTi) tubes configured in a versatile multimode heat exchange architecture, which allows the harnessing of both high delivered cooling power and large temperature spans. Our system shows that elastocaloric cooling, which only emerged 8 years ago, is a promising direction for commercializing caloric cooling.

From air-conditioning to food preservation, refrigeration is a ubiquitous and indispensable part of modern life, and the vapor compression cycle that dominates all cooling technologies accounts for ~20% of global energy consumption. An exacerbating complication is the increasing reliance on vapor compression–based heat pumps that is driven by the concerted community effort to "decarbonize and electrify" fossil fuel–based furnaces for space heating and industrial applications (*1*), a trend magnified recently by the acute supply problem of natural gas and crude oil in Europe (*2*).

Hydrofluorocarbons used as refrigerants in vapor compression systems are super-polluting greenhouse gases—their global warming potential (GWP) is hundreds to thousands of times worse than that of CO_2, and their leakage is projected to contribute to >11% of global CO_2 emission by the year 2050 (*3*). A recent directive by the Biden administration calls for an 85% phase-down of hydrofluorocarbons in the US in the next 15 years, but there are currently no ideal alternative low-GWP refrigerants, because of the trade-off between their environmental impact and safety concerns, such as flammability (*4*).

Among zero-GWP, nonflammable, and nontoxic alternatives to vapor compression, caloric cooling (and heating) is a potential frontrunner. Caloric cooling uses field-driven phase transitions in solids and has intrinsically high energy conversion efficiency (*5*–*7*). Caloric materials include magnetocaloric (*8*), electrocaloric (*9*), and mechanocaloric materials (*10*–*16*). Magnetocaloric materials and devices have been explored for almost 50 years, with kilowatt-range cooling power refrigeration systems being demonstrated (*17*). Their high magnetic field requirement (>1 T) has, however, hampered commercialization of the magnetocaloric cooling technology owing to the rising criticality of the rare earth element–based permanent magnets. Electrocaloric devices have recently shown temperature spans as large as ~10 K in regenerative and multistage configurations (*18*, *19*) and have demonstrated application potential for cooling electronic devices that require localized temperature control.

Mechanocaloric cooling is driven by mechanical stress, and—reflecting the diversity of ways in which materials can be subjected to stress (*20*)—it can be induced by uniaxial stress (elastocaloric) (*10*–*12*), torsional stress (twistocaloric) (*13*), bending stress (flexocaloric) (*14*), or hydrostatic pressure (barocaloric) (*15*, *16*). Owing to large intrinsic materials temperature lifts of elastocaloric materials (*21*), elastocaloric cooling devices represent some of the most promising caloric technologies (*22*–*26*). Compared with their magnetocaloric and electrocaloric counterparts, which usually require large fields that are expensive to apply, there are practical actuators that can be implemented for driving elastocaloric devices. Most elastocaloric devices are based on superelastic shape-memory alloys (*27*–*36*), although prototypes using polymers have also been reported (*37*, *38*). The largest adiabatic temperature change ($\Delta T_{adiabatic}$) recorded from direct measurements on caloric materials under a practical range of fields are from elastocaloric materials including NiMnTi and NiTi, which displayed $\Delta T_{adiabatic}$ of 31.5 and 38.5 K, respectively (*21*, *39*).

Some of the successful caloric cooling prototypes are based on the active regeneration scheme, which is used to boost the temperature span across the devices from limited adiabatic temperature changes of caloric materials, in particular magnetocaloric and electrocaloric materials (*18*, *40*). Because the adiabatic temperature change of elastocaloric materials can exceed 20 K, nonregenerative single-stage elastocaloric systems are viable for applications such as air conditioners or refrigerators. Single-stage operation is characterized by a uniform temperature inside materials at any stage of each cycle, which is facilitated by a large utilization factor (percentage of heat exchange medium that is displaced during each cycle), allowing most of the caloric effect to be directly transferred from the refrigerant materials to the heat sink and heat source. Examples of single-stage elastocaloric systems are tensile prototypes using TiNi sheets or wires based on a solid-to-solid contact heat exchange mechanism (*27*–*29*, *35*), compressive prototypes using NiTi tubes incorporating various heat exchange fluid schemes (*32*, *41*), and the biaxial stretching prototype using rubber membranes (*37*).

To enhance the temperature span of elastocaloric prototypes, cascading multiple single-stage schemes have been experimentally demonstrated in a four-stage device with a 27 K temperature span (*31*) and a three-stage fluid-based device using NiTi wires (*30*). Active regeneration has also been successfully implemented for elastocaloric cooling with a temperature span close to 20 K (*33*, *34*) and 31.3 K with 60 W cooling power recently (*36*). Unlike single-stage operation, active regeneration is characterized by a large temperature gradient across the caloric material at any phase of each cycle, where a small utilization factor is essential to maintain the enhanced temperature gradient. Nonetheless, the enhanced temperature span in active regenerators comes at a cost of a considerable portion of the sensible heat remaining inside the regenerator in order to establish and sustain the temperature span.

Design of the multimode elastocaloric cooling system

To capture the best aspects of the active regeneration cycle and the large utilization operation in a single practical prototype, we developed a multimode elastocaloric cooling system leveraging the large temperature span of the active regeneration mode (Fig. 1A) and the efficient cooling of the maximum

[1]Department of Refrigeration and Cryogenic Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, People's Republic of China. [2]Center for Environmental Energy Engineering, Department of Mechanical Engineering, University of Maryland, College Park, MD 20742, USA. [3]Department of Materials Science and Engineering, University of Maryland, College Park, MD 20742, USA. [4]Key Laboratory of Aerospace Materials and Performance (Ministry of Education), School of Materials Science and Engineering, Beihang University, Beijing 100191, People's Republic of China. [5]Zhongfa Aviation Institute of Beihang University, Hangzhou, Zhejiang 310023, People's Republic of China. [6]Tianmushan Laboratory (Zhejiang Provincial Laboratory for Aviation), Hangzhou, Zhejiang 310023, People's Republic of China. [7]Maryland Quantum Materials Center, Department of Physics, University of Maryland, College Park, MD 20742, USA.
*Corresponding author. Email: takeuchi@umd.edu
†These authors contributed equally to this work.
‡Present address: DTU Wind and Energy Systems, 4000 Roskilde, Denmark.

utilization mode (Fig. 1B). In the prototype, the two modes can be easily switched from one to the other by controlling the operation sequences of valves in the heat exchange fluid network (figs. S1 and S2).

The deviatoric stress state associated with the phase transformation in elastocaloric materials needs to be taken into account to design the field actuation and the cycle lifetime in elastocaloric systems. We have chosen the uniaxial stress—the simplest deviatoric stress state—realized by the compressive or tensile actuation to implement in our elastocaloric cooling system. Although both compressive and tensile loadings can produce similar elastocaloric effects (*42*), the compressive actuation can hinder the advancing of microcracks, thereby substantially extending the lifetime of the elastocaloric materials (*43*–*47*), and elastocaloric materials have been reported to endure >10 million cycles in compression modes (*32*). With 0.1 Hz operation for a usage modality of 12 hours per day and 180 days per year, 10 million cycles corresponds to 12.8 years, beyond the standard service life of a commercial appliance. Thus, we have enlisted compressive actuation in our multimode elastocaloric cooling system. The ideal geometry of the elastocaloric materials not only exhibits robust mechanical integrity under cyclic compression but is also conducive to heat exchange between the material and the heat-transfer fluid (distilled water). To this end, we have designed our system around readily available NiTi tubes (manufactured by Confluent Medical; properties detailed in fig. S3), which can be configured into a bundle for relative ease of compression as well as for flowing heat exchange fluid with minor hydraulic losses in the uniaxial direction (*48*). Our elastocaloric system consisting of four bundles of NiTi tubes can deliver up to 260 W of useful cooling (measured against heat load) operating in the maximum utilization mode, one of the largest reported cooling powers for elastocaloric devices to date, as well as a temperature span of 22.5 K in the active regeneration mode. With such high-performance characteristics, our elastocaloric system rivals some of the best reported magnetocaloric systems (*49*).

Nineteen NiTi tubes are packed in a hexagonal bundle configuration, surrounded by a layer of stainless-steel tubes that prevent the NiTi tubes from direct contact with the bundle holder sleeve. The holder is necessary to circumvent buckling of the tubes, and the stainless-steel tubes work to minimize the friction that the outer NiTi tubes experience during compression (Fig. 1C). Each NiTi tube bundle is compressed in the axial direction by two loading heads. The loading heads also serve as a fluid manifold to evenly disperse the heat exchange fluid into the 19 NiTi tubes. The profile of the flow channel inside the loading heads has been optimized on the basis of the simulated streamline. To enhance the heat exchange efficiency between the fluid and the NiTi tubes, triangle polymer inserts are placed inside each NiTi tube to block ~40% of the cross-sectional area (fig. S4).

Our multimode elastocaloric cooling system consists of two sets of two (paired) bundles of NiTi tubes. For simplicity, we only show one

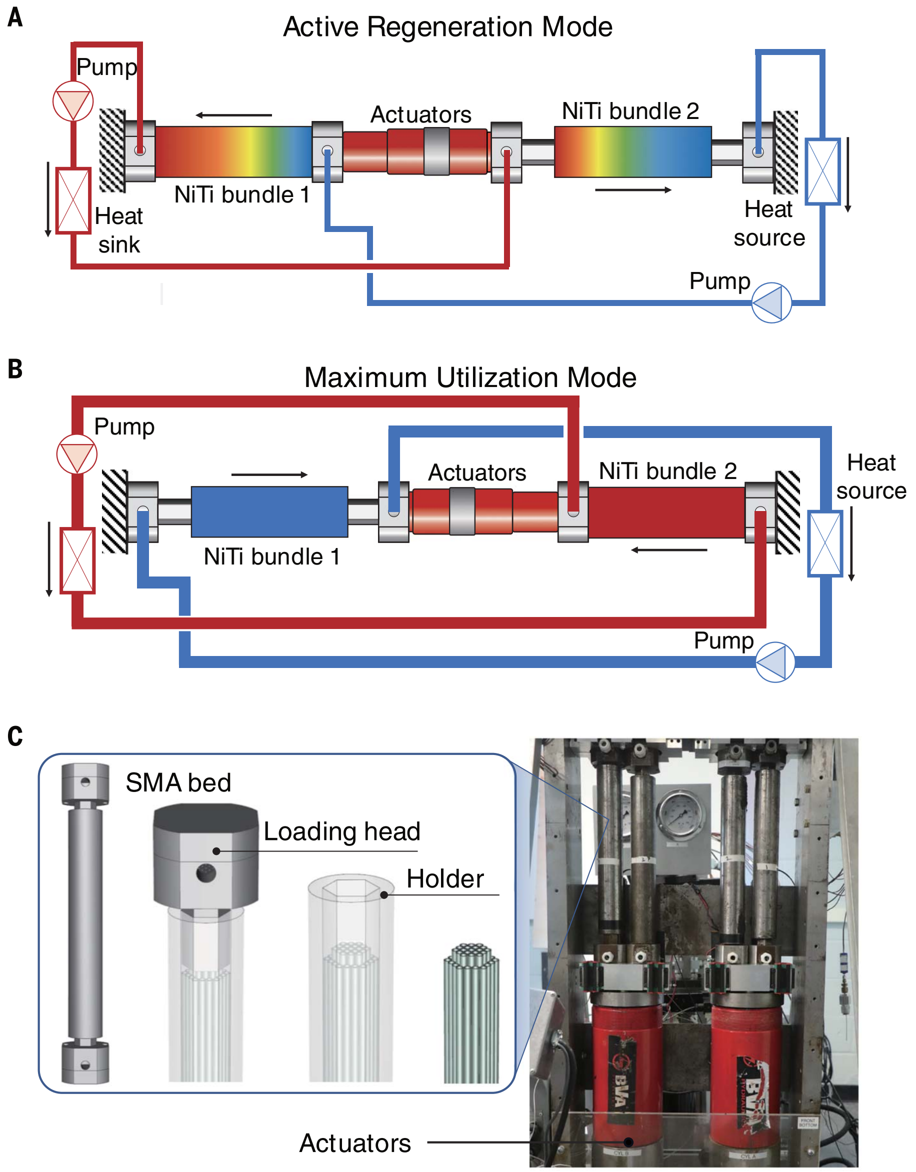


Fig. 1. System design schematics and operating principles of the multimode elastocaloric cooling system. (**A**) Active regeneration mode for achieving a large temperature span, which develops across each NiTi bundle located inside a tube assembly. The snapshot represents the phase of the cycle where bundle 1 is being compressed, and heat exchange fluid flows from the cold end of the bundle toward the hot end, while, at the same time, bundle 2 is being unloaded, and fluid (distilled water) flows from the hot end of the bundle to the cold end. (**B**) Maximum utilization mode (which provides enhanced cooling power, as indicated by the complete extraction of elastocaloric effect in both bundles with uniform temperature) at the phase of the cycle when bundle 2 is being compressed, and the bundle is connected to the heat sink fluid circuit, while the other bundle is unloaded and connected to the heat source fluid circuit. For both modes, compression work done to one bundle by the actuator in the middle is always assisted by synchronized unloading in the other bundle, achieving work recovery. The two modes can be switched by changing valve sequences (figs. S1 and S2). (**C**) A photo of the central part of the multimode elastocaloric cooling system, which stands at 0.3 m by 0.8 m with a height of 1.3 m, with detailed schematics of the NiTi tube assembly. From left to right: the completed assembly with loading heads on both ends; the loading head compresses on NiTi tubes; the complete bundle inside the holder; and 19 NiTi tubes in a bundle surrounded by 18 shorter, flexible stainless tubes, shielding NiTi tubes from coming in direct contact with holder wall. Heat exchange fluid flows inside each NiTi tube.

pair of tube bundles, denoted as bundle 1 and bundle 2 (Fig. 1, A and B). The two paired bundles are operated a half-cycle out of phase with each other to achieve work recovery, where compression of one bundle is always assisted by the unloading force of the other, to achieve high-efficiency operation.

In both the active regeneration mode and the maximum utilization mode, the two paired bundles of NiTi tubes are cycled in four major steps: (i) the actuator compresses one bundle and unloads the other bundle adiabatically without flowing any fluid; (ii) the first bundle rejects heat to the heat exchange fluid, while the second bundle absorbs heat from the fluid and provides cooling; (iii) the actuator unloads the first bundle and compresses the second bundle adiabatically; and finally, (iv) the first bundle cools the heat exchange fluid, while the second bundle rejects heat to the fluid.

The major difference between the two modes is in the details of the heat exchange processes. When a large temperature span is desired (for instance, for air dehumidification or for heat pump applications), the heat exchange fluid network is operated in such a way that each bundle of NiTi tubes is simultaneously connected to the heat sink and the heat source, to which cooling is delivered (Fig. 1A). This mode is for active regeneration, where a unidirectional fluid network is implemented to minimize dead fluid volume (*50*). In this mode, the displaced fluid volume is less than the fluid volume inside the NiTi tubes, and only a portion of the latent heat from the NiTi tubes is pumped to the heat sink and the heat source. Most of the latent heat remains inside the NiTi tubes to build up the fluid temperature before the fluid exits the NiTi tubes during heat rejection and lowers the fluid temperature during the cooling process, and a temperature gradient along the flow direction can be sustained to establish a large temperature span (fig. S1), as has been applied in magnetocaloric cooling systems for decades (*51*).

The maximum utilization mode can be implemented when a relatively large cooling power is required, for example, during the initial turn-on process for an air conditioner or a refrigerator. The maximum utilization mode features at least a 10-fold increase in displaced fluid volume compared with the active regeneration mode, leading to a homogeneous temperature distribution inside the NiTi tubes, and most of the latent heat can be extracted and used in the heat sink and the heat source. To fully displace the remnant fluid inside the NiTi tubes, the valves are switched so that the first bundle is only connected to the heat source when the NiTi tubes are unloaded, while the second bundle is connected to the heat sink after it is loaded (Fig. 1B). In the maximum utilization mode, heat recovery by exchanging heat between the two paired bundles can be applied to precool the first bundle before

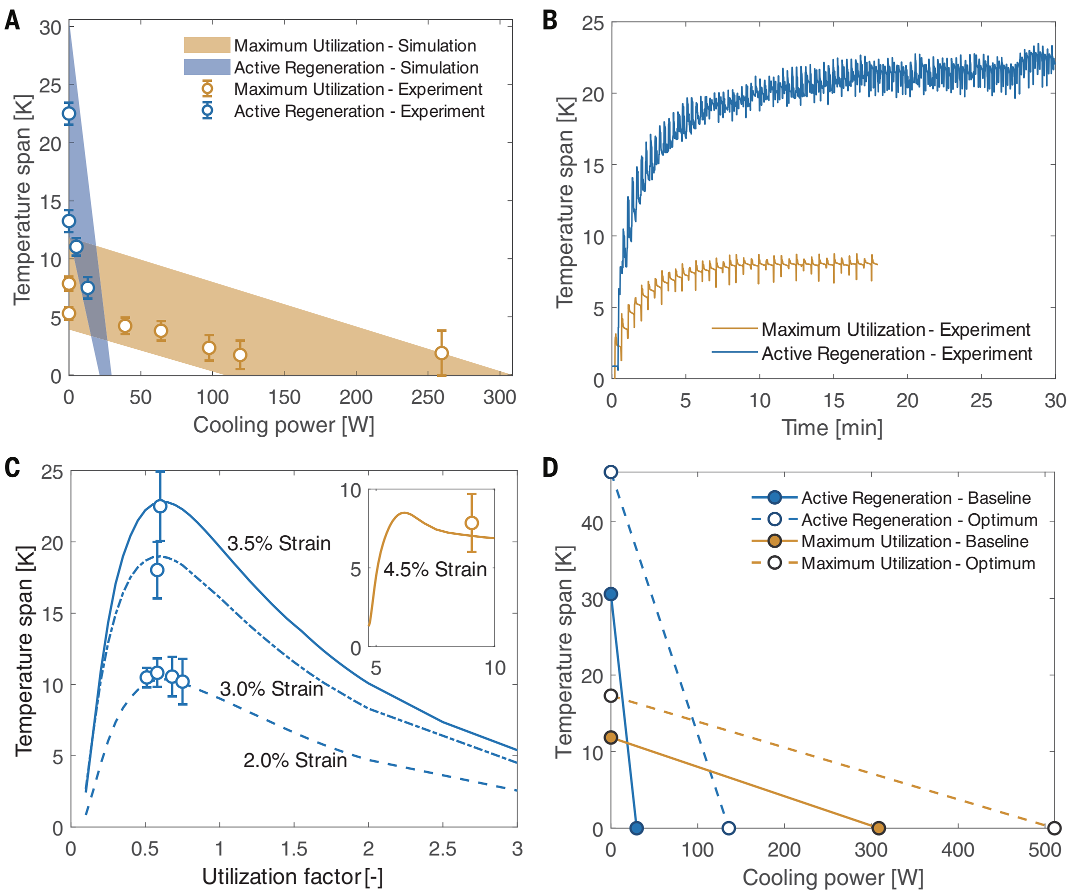


Fig. 2. Performance of the multi-mode elastocaloric cooling system. (**A**) Temperature span versus cooling power for the active regeneration mode (blue) and the maximum utilization mode (orange). The data points are experimentally measured values (error bars represent ±SD of fluid temperature during steady-state operation, and the number of repeated cycles during steady-state *N* equals 10), and the shaded regions are ranges observed in simulations taking into account various heat leak coefficients (table S1) and various operating conditions. For the active regeneration mode and maximum utilization mode, the NiTi tubes are compressed to 3.5 and 4.5% for loading, corresponding to a strain rate of 8.75×10^{-3} s^{-1} and 7.5×10^{-3} s^{-1}, respectively. The active regeneration mode cannot be operated at strains higher than 3.5% owing to the nonuniform strain distribution (see section 4 of the materials and methods section). (**B**) The evolution of temperature span as a function of time. For the maximum utilization mode, the saturation temperature span of 8 K is reached after 10 min and 20 cycles. The active regeneration mode takes longer to saturate, but the final temperature span is much larger, ~22.5 K. Detailed temperatures of heat sink and heat source are available in fig. S6. (**C**) Temperature span as a function of the utilization factor. The blue curves are simulation data at three different strains for the active regeneration mode, where the temperature span peaks sharply for all three tested strains at a utilization factor of 0.6, accompanied by the experimental data points. The orange curve (simulation) and the experimental data point are for the maximum utilization mode. (**D**) Expected performance enhancement upon implementing various inserts inside NiTi tubes for promoting heat exchange. Blue solid and dashed lines are the performance of active regeneration mode with baseline insertion and optimized insertion, respectively. Orange solid and dashed lines are the performance of maximum utilization mode with baseline insertion and optimized insertion, respectively.

unloading and preheat the second bundle before loading, which adds the additional two steps to the cycle (fig. S2) (*52*).

Performance of the multimode elastocaloric cooling system

Any refrigeration device has a trade-off between the cooling power and the temperature span. This is because the total sum of the delivered cooling power and the cooling effect that sustains the temperature span remains constant according to the first law of thermodynamics. Such a trade-off is captured in the performance Pareto front curve on the cooling power–ΔT diagram. In our multimode elastocaloric cooling system, the maximum utilization mode has delivered 260 W at a 1.9 K temperature span at one end of the performance Pareto front (Fig. 2A). This wattage is one of the largest delivered cooling powers (measured against real heat load of an electric heater) of any elastocaloric cooling system to date, and it can be boosted by simply increasing the amount of NiTi material or by using a different material with a larger latent heat.

On the other end of the performance Pareto front, in the active regeneration mode, we achieved a ΔT as large as 22.5 K (Fig. 2B). This 22.5 K temperature span is larger than the material's adiabatic temperature change of 11 K at a strain level of 3.5% (fig. S5), resulting in a regeneration factor of 2.0 (*18*), which illustrates the effectiveness of an active regenerator. By combining the maximum utilization mode and the active regeneration mode, the performance domain in the cooling power–ΔT diagram can be expanded, superseding the best performance of each respective mode alone.

The key controlling factor that bridges the two modes is the utilization factor, which determines the percentage of heat exchange fluid that is displaced during each cycle (eqs. S1 and S2). The optimum utilization factor controls the competition between the latent heat to be extracted from the NiTi tubes and the elastocaloric effect that maintains the temperature gradient in both the active regenerator and the passive regenerator (in the maximum utilization mode) (Fig. 2C). Such strong dependence of the overall performance of our devices on the utilization factor is consistent with previous reports on elastocaloric and magnetocaloric active regenerators (*33*, *53*). Our multimode elastocaloric cooling system has successfully extended the range of the optimum utilization factor from less than one in the active regeneration mode to more than six in the maximum utilization mode, adding flexibility in operations and thus applications of the elastocaloric cooling system.

We can introduce several measures to further improve the performance of the current system. One is to further reduce the fluid-to-solid

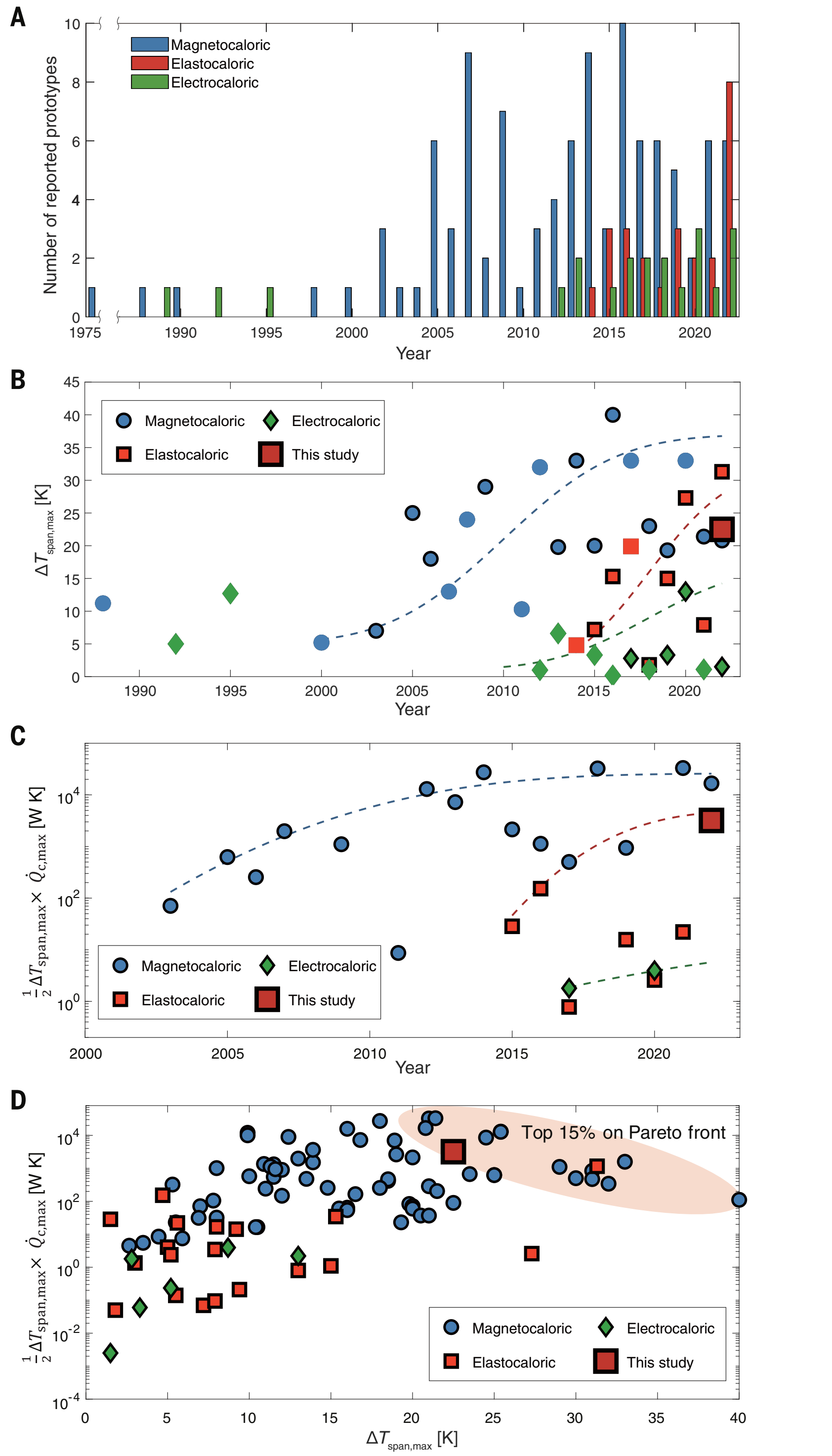


Fig. 3. Comparison of reported elastocaloric, magnetocaloric, and electrocaloric cooling prototypes and their performance. (**A**) The number of publications for magnetocaloric (blue), elastocaloric (red), and electrocaloric (green) prototypes. (**B**) The maximum reported performance in each year, $\Delta T_{span,max}$. Markers with black outlines are prototypes for which both temperature span and cooling power were reported, whereas markers without outlines indicate prototypes for which only temperature span was reported. Dashed lines represent the trend in recent years. The temperature span and cooling power of magnetocaloric systems using an electromagnet or helium are excluded because they are not commercially practical, but they were included in the number of reported prototypes. (**C**) The maximum reported values of the figure of merit (FOM) in that year, $\frac{1}{2}\Delta T_{span,max} \times \dot{Q}_{c,max}$, which equals the area under the Pareto front curve in ΔT–cooling power plots (fig. S7). (**D**) Performance of all reported caloric devices in FOM–$\Delta T_{span,max}$ plot. The complete publication list is given in tables S2 to S4.

thermal mass ratio inside the NiTi tubes by varying the geometries of the inserts (fig. S4). Improved heat transfer is expected in NiTi tubes with an increased blocked central flow area, because the heat transfer in the boundary layer becomes more efficient and the thermal inertia of the fluid is reduced. Our numerical model shows that applying circular-rod inserts with 82% blockage of the flow channel in each NiTi tube could increase the cooling power by 65% at zero temperature span in the maximum utilization mode, as compared with the current triangle-shaped inserts (Fig. 2D). Also, increasing the cycle frequency of the active regeneration mode can directly enhance the cooling power. For instance, by going from the current frequency of 0.071 Hz to 0.125 Hz, the cooling power can be enhanced by a factor of 1.2.

Given the continuing call to phase down high-GWP hydrofluorocarbons, and the pace with which advances are made in various caloric materials and devices in recent years, we are compelled to survey and juxtapose the main caloric technologies at this juncture (Fig. 3). Figure 3A plots the number of reported magnetocaloric, elastocaloric, and electrocaloric prototypes (regenerators and systems) publications versus year (full list of publications is given in tables S2 to S4). Development activities were dominated by magnetocaloric systems until recently, but there has been a steady increase in elastocaloric and electrocaloric prototype demonstrations. Other emerging caloric technologies are also on the horizon, for example, barocaloric and ionocaloric cooling (*15*, *54*), but without development of devices based on them with heat exchangers that produce cooling power.

Among various performance metrics, we have elected to use temperature span ΔT (Fig. 3B) as one of the key indicators of technological advances over the years. We plot the reported maximum temperature spans ΔT measured across devices and regenerators for various reported caloric prototypes (Fig. 3B). The ΔT of recent magnetocaloric devices have hovered consistently around 35 K, and they reflect the progress in rotary and other regenerators, which are effective in boosting adiabatic temperature change of Gd of 3 to 4 K (and LaFeSiH of 5 to 8 K) at 2 T (*55*). The first elastocaloric cooling device, reported in 2014 by Schmidt *et al.*, had a ΔT of 5 K (*56*), and since then, there has been a remarkable surge in the reported ΔT of elastocaloric systems, and the largest reported ΔT from an elastocaloric regenerator to date is 31.3 K (rivaling ΔT of many reported magnetocaloric devices) (*36*). This upward trend reflects the fact that (i) the starting adiabatic temperature change of elastocaloric materials such as NiTi is much higher than those of typical magnetocaloric materials and (ii) researchers have been able to adopt the active regeneration concept from magnetocaloric systems (*33*). Electrocaloric regenerators also have reported ΔT as large as 13 K measured across regenerators (*18*).

As seen in the design and operation of our multimode cooling system, cooling power and temperature span are equally important in practical systems, and thus we have opted to use the product of the maximum reported ΔT and the maximum reported cooling power (watts) ($\times^1/_2$, translating to the area under the curve in ΔT–cooling power plots, e.g., fig. S7 and eq. S16) as another practical measure of the overall performance of reported devices (Fig. 3C). In this comparison as well, the rapid rise in the performance of the best reported elastocaloric systems in the past few years is evident, and they are beginning to rival magnetocaloric prototypes. In fact, the performance of our multimode elastocaloric cooling system is among the top 15% in this metric and in ΔT among all hitherto reported caloric cooling systems in the past four decades, as reflected by the performance that approaches the Pareto front (Fig. 3D).

Operating and switching multiple bundles of NiTi tubes to operate in different modes can facilitate high-efficiency performance in a wide range of working conditions. Efficiency—or, alternatively, the coefficient of performance (COP)—should be compared at the same temperatures and under the same conditions (*20*). All previously reported COP had excluded the efficiency of work recovery and the efficiency of the actuator, which are important factors. Our multimode elastocaloric cooling system achieved a modest system-level COP ("plugged-into-electrical-outlet" COP) of 1.03 at a temperature span of 1.9 K (see section 6 of the materials and methods section in the supplementary materials). This value may be further improved to the regenerator-level COP of 6.85 if more-efficient actuators and kinetic energy recovery are applied.

The multimode elastocaloric cooling concept can be extended to cascading multiple bundles of NiTi tubes, so that variations in the modes can be implemented. For instance, cascading two bundles at active regeneration mode or maximum utilization mode is expected to help fill the performance gap in Fig. 2A and can realize a smooth transition between the two modes. We believe that such versatility is the key to the successful commercialization of this zero-GWP technology. We also expect to be able to implement Cu-based elastocaloric materials, which require much smaller stress than NiTi, enabling more-efficient system operation with smaller actuators (*57*).

REFERENCES AND NOTES

1. X. Shen, P. Liu, Y. Qiu, A. Patwardhan, P. Vaishnav, *Nat. Energy* **6**, 30–37 (2021).
2. Borders and pipelines. *Nat. Energy* **7**, 213 (2022).
3. D. Coulomb, J.-L. Dupont, V. Morlet, "The impact of the refrigeration sector on climate change, 35th Informative Note on Refrigeration Technologies" (International Institute of Refrigeration, 2017); https://iifiir.org/en/fridoc/the-impact-of-the-refrigeration-sector-on-climate-change-141135.
4. M. O. McLinden, C. J. Seeton, A. Pearson, *Science* **370**, 791–796 (2020).
5. S. Fähler *et al.*, *Adv. Eng. Mater.* **14**, 10–19 (2012).
6. X. Moya, S. Kar-Narayan, N. D. Mathur, *Nat. Mater.* **13**, 439–450 (2014).
7. I. Takeuchi, K. Sandeman, *Phys. Today* **68**, 48–54 (2015).
8. A. Kitanovski, *Adv. Energy Mater.* **10**, 1903741 (2020).
9. X. Qian *et al.*, *Nature* **600**, 664–669 (2021).
10. E. Bonnot, R. Romero, L. Mañosa, E. Vives, A. Planes, *Phys. Rev. Lett.* **100**, 125901 (2008).
11. C. Chluba *et al.*, *Science* **348**, 1004–1007 (2015).
12. H. Hou *et al.*, *Science* **366**, 1116–1121 (2019).
13. R. Wang *et al.*, *Science* **366**, 216–221 (2019).
14. A. Czernuszewicz, L. Griffith, J. Slaughter, V. Pecharsky, *Appl. Mater. Today* **19**, 100557 (2020).
15. B. Li *et al.*, *Nature* **567**, 506–510 (2019).
16. L. Mañosa *et al.*, *Nat. Mater.* **9**, 478–481 (2010).
17. S. Jacobs *et al.*, *Int. J. Refrig.* **37**, 84–91 (2014).
18. A. Torelló *et al.*, *Science* **370**, 125–129 (2020).
19. Y. Wang *et al.*, *Science* **370**, 129–133 (2020).
20. H. Hou, S. Qian, I. Takeuchi, *Nat. Rev. Mater.* **7**, 633–652 (2022).
21. D. Cong *et al.*, *Phys. Rev. Lett.* **122**, 255703 (2019).
22. X. Moya, N. D. Mathur, *Science* **370**, 797–803 (2020).
23. W. Goetzler, R. Zogg, J. Young, C. Johnson, "Energy savings potential and RDD opportunities for non-vapor-compression HVAC technologies" (Tech. Rep. DOE/EE-1021, US Department of Energy, Office of Energy Efficiency and Renewable Energy, Building Technologies, Office, 2014); https://doi.org/10.2172/1220817.
24. L. Mañosa, A. Planes, *Appl. Phys. Lett.* **116**, 050501 (2020).
25. S. Qian *et al.*, *Int. J. Refrig.* **62**, 177–192 (2016).
26. S. Qian *et al.*, *Int. J. Refrig.* **64**, 1–19 (2016).
27. M. Schmidt, A. Schütze, S. Seelecke, *Int. J. Refrig.* **54**, 88–97 (2015).
28. H. Ossmer *et al.*, *Smart Mater. Struct.* **25**, 085037 (2016).
29. F. Brüederlin, H. Ossmer, F. Wendler, S. Miyazaki, M. Kohl, *J. Phys. D Appl. Phys.* **50**, 424003 (2017).
30. R. Snodgrass, D. Erickson, *Sci. Rep.* **9**, 18532 (2019).
31. F. Brüederlin, thesis, Karlsruher Institut für Technologie (KIT) (2020).
32. N. Bachmann *et al.*, *Commun. Phys.* **4**, 194 (2021).
33. J. Tušek *et al.*, *Nat. Energy* **1**, 16134 (2016).
34. K. Engelbrecht *et al.*, *J. Phys. D Appl. Phys.* **50**, 424006 (2017).
35. Y. Chen, Y. Wang, W. Sun, S. Qian, J. Liu, *Innovation (Camb.)* **3**, 100205 (2022).
36. Ž. Ahčin *et al.*, *Joule* **6**, 2338–2357 (2022).
37. F. Greibich *et al.*, *Nat. Energy* **6**, 260–267 (2021).
38. D. W. Gerlach, *Proc. ASME* **6**, 289–293 (2009).
39. L. Ding *et al.*, *Acta Mater.* **218**, 117200 (2021).
40. G. V. Brown, *J. Appl. Phys.* **47**, 3673–3680 (1976).
41. S. Qian *et al.*, "Experimental evaluation of compressive elastocaloric cooling system," 16th International Refrigeration and Air Conditioning Conference at Purdue, West Lafayette, IN, 11 to 14 July 2016, paper 1726.
42. Y. Wu, E. Ertekin, H. Sehitoglu, *Acta Mater.* **135**, 158–176 (2017).
43. H. Hou *et al.*, *MRS Bull.* **43**, 285–290 (2018).
44. J. Chen, K. Zhang, Q. Kan, H. Yin, Q. Sun, *Appl. Phys. Lett.* **115**, 093902 (2019).
45. P. Hua, M. Xia, Y. Onuki, Q. Sun, *Nat. Nanotechnol.* **16**, 409–413 (2021).
46. L. Porenta *et al.*, *Appl. Mater. Today* **20**, 100712 (2020).
47. L. Ianniciello, K. Bartholomé, A. Fitger, K. Engelbrecht, *Appl. Therm. Eng.* **202**, 117838 (2022).
48. S. Qian *et al.*, *Sci. Technol. Built Environ.* **22**, 500–506 (2016).
49. D. Eriksen *et al.*, *Int. J. Refrig.* **58**, 14–21 (2015).
50. C. Zimm *et al.*, *Int. J. Refrig.* **29**, 1302–1306 (2006).
51. A. Kitanovski *et al.*, *Magnetocaloric Energy Conversion: From Theory to Applications* (Springer International Publishing, 2015).
52. S. Qian, J. Yu, G. Yan, *Renew. Sustain. Energy Rev.* **69**, 535–550 (2017).
53. M. Masche, J. Liang, S. Dall'Olio, K. Engelbrecht, C. R. H. Bahl, *Appl. Therm. Eng.* **199**, 117569 (2021).
54. D. Lilley, R. Prasher, *Science* **378**, 1344–1348 (2022).
55. V. Franco *et al.*, *Prog. Mater. Sci.* **93**, 112–232 (2018).
56. M. Schmidt, A. Schütze, S. Seelecke, *Proc. ASME* **2**, V002T04A013 (2014).
57. S. Qian *et al.*, *Philos. Trans. R. Soc. London Ser. A* **374**, 20150309 (2016).

ACKNOWLEDGMENTS

We thank Y. Wu, L. Yuan, P. Gill, T. Yamazaki, and A. Santos for assistance and discussions. **Funding:** This work was supported by the US Department of Energy under DE-EE0009159 and the Advanced Research Projects Agency-Energy (ARPA-e) under ARPA-e DEAR0000131. S.Q. acknowledges the support of the National Natural Science Foundation of China (grant 51976149),

the Foundation for Innovative Research Groups of the National Natural Science Foundation of China (grant 51721004), and the Young Elite Scientist Sponsorship Program by CAST (grant 2019QNRC001) for numerical simulations. H.H. acknowledges the support of the National Natural Science Foundation of China (grant 12002013), the National Key Research and Development Program of China (grant 2022YFB3800052), Zhejiang Natural Science Foundation (grant LZ23E010004), the Fundamental Research Funds for the Central Universities (grant 501LKQB2020105028), and the Opening Fund of State Key Laboratory of Nonlinear Mechanics for materials characterizations. **Author contributions:** I.T., R.R., and Y.H. initiated and supervised the research. S.Q., D.C., J.M., R.R., Y.H., and I.T. designed and constructed the elastocaloric systems at the University of Maryland. D.C., S.Q., J.M., B.L., and H.M. operated and took measurements on the elastocaloric systems. D.C. and S.Q. performed simulations. S.Q., D.C., and H.H. evaluated the materials properties. I.T., S.Q., D.C., and H.H. wrote the paper with substantial input from the other authors. All authors contributed to the discussion of the results. **Competing interests:** I.T. is a founder of Maryland Energy and Sensor Technologies, LLC, a company that works on elastocaloric technologies. The authors have been granted US patents 10119059, 10018385, and 10823465 related to this work. **Data and materials availability:** All data are available in the manuscript or the supplementary materials. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adg7043
Materials and Methods
Figs. S1 to S13
Tables S1 to S4
References (*58–200*)

Submitted 6 February 2023; accepted 13 April 2023
10.1126/science.adg7043

CATALYSIS

Disentangling the activity-selectivity trade-off in catalytic conversion of syngas to light olefins

Feng Jiao[1,2]†, Bing Bai[1,2]†, Gen Li[1,2]†, Xiulian Pan[1,2]*, Yihan Ye[1,2], Shengcheng Qu[1], Changqi Xu[1], Jianping Xiao[1,2], Zhenghao Jia[1], Wei Liu[1], Tao Peng[1,3], Yilun Ding[1,2], Cheng Liu[1,2], Jinjing Li[1], Xinhe Bao[1,2,3]*

Breaking the trade-off between activity and selectivity has been a long-standing challenge in the field of catalysis. We demonstrate the importance of disentangling the target reaction from the secondary reactions for the case of direct syngas conversion to light olefins by incorporating germanium-substituted AlPO-18 within the framework of the metal oxide–zeolite (OXZEO) catalyst concept. The attenuated strength of the catalytically active Brønsted acid sites allows enhancing the targeted carbon-carbon coupling of ketene intermediates to form olefins by increasing the active site density while inhibiting secondary reactions that consume the olefins. Thus, a light-olefins selectivity of 83% among hydrocarbons and carbon monoxide conversion of 85% were obtained simultaneously, leading to an unprecedented light-olefins yield of 48% versus current reported light-olefins yields of ≤27%.

Achieving both high activity and high selectivity is a long-standing challenge for selective reactions such as the hydrogenation of CO in direct syngas conversion (*1*–*8*). Fischer-Tropsch synthesis (FTS), as a key process for syngas conversion using Fe-based (*8*–*10*) and Co-based (*11*–*13*) catalysts, allows direct synthesis of light olefins (two to four carbon atoms, $C_2^=–C_4^=$), named as Fischer-Tropsch-to-olefins (FTTO). Substantial progress has been made in the past decade (table S1). For example, a 61% light-olefins selectivity among hydrocarbons at 0.5 to 1% CO conversion and 52% selectivity at 88% CO conversion had been obtained, giving 26% yield of light olefins over Fe/carbon nanofiber catalysts (*8*). Hydrophobically modified FeMn catalysts gave 63% selectivity for all olefins and ~27% for light olefins at 56% CO conversion, amounting to ~14% yield of light olefins (*9*). A light-olefins selectivity of 61% among hydrocarbons was achieved at 32% CO conversion over Mn-promoted Co_2C catalysts, giving 10% $C_2^=–C_4^=$ yield (*11*). More recently, CO conversion was enhanced to 64%, with 71% light-olefins selectivity among hydrocarbons when CoMn carbide was physically mixed with hydrophobic polydivinylbenzene, giving 24% yield of light olefins (*13*).

In addition, bifunctional metal oxide–zeolite (OXZEO) catalysts were recently developed for direct syngas conversion to light olefins (*6*, *7*, *14*–*16*). By separating the key steps of CO and H_2 activation from the subsequent C–C coupling of the ketene intermediates (CH_2CO, which is a carrier of CH_2) on the active sites of metal oxides and zeolites (or zeotypes), respectively, the product selectivity can be modulated through the topology of zeolites (or zeotypes), and light-olefins selectivity higher than 80% was obtained (*6*, *7*, *16*). Recently, the technology was successfully demonstrated in a pilot plant with 1000 ton/year light-olefins output in Shaanxi, China (*7*, *17*). Extensive efforts have been made to optimize the catalysts by screening a wide range of metal oxide and zeolite (zeotype) combinations as well as reaction conditions (*7*, *16*, *18*–*24*). As a result, the catalytic performance has been greatly improved; CO conversion could be increased from 17 to 60% with light-olefins selectivity being maintained at 75%, and the yield of light olefins improved from 8 to 27% (*16*, *21*, *24*). However, further improvement of OXZEO and FTTO processes appears to be stalled by an activity-selectivity trade-off (Fig. 1A, dashed line, and table S1).

The mechanistic origin of the activity-selectivity trade-off can be traced to the catalytic sites for both the target and side reactions on the catalyst. Taking the silicoaluminophosphate (SAPO)–type zeotypes in the OXZEO syngas conversion concept as an example, the Si–OH–Al Brønsted acid sites are the active sites, catalyzing C–C coupling of the ketene intermediates to form light olefins as products (*7*, *18*), which consequently shifts the thermodynamically limited step of syngas conversion to ketene (*7*, *25*). A high selectivity of light olefins has been achieved by weakening the acidity through reduction of the Si content (*7*, *16*, *18*). However, this reduces simultaneously the density of the Si–OH–Al acid sites and hence the rate of intermediate consumption and ultimately CO conversion. CO conversion can be facilitated with a higher acid site density by increasing the Si content of the SAPO zeotypes. But this comes at the expense of olefins selectivity because the Si–OH–Al Brønsted acid sites are

[1]State Key Laboratory of Catalysis, 2011-Collaborative Innovation Center of Chemistry for Energy Materials, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China. [2]University of Chinese Academy of Sciences, Beijing 100049, China. [3]University of Science and Technology of China, Hefei 230026, China.
***Corresponding author. Email: panxl@dicp.ac.cn (X.P.); xhbao@dicp.ac.cn (X.B.)**
†These authors contributed equally to this work.

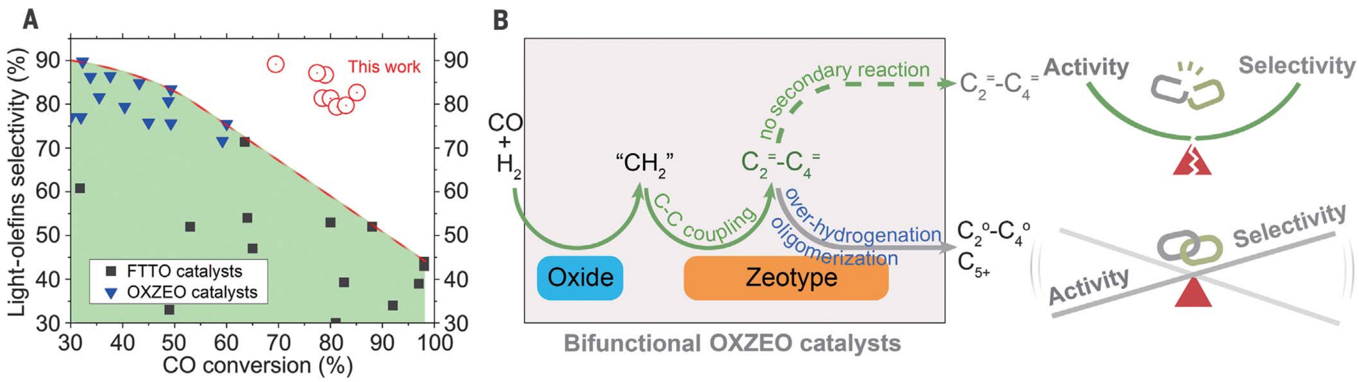


Fig. 1. The activity-selectivity trade-off in the OXZEO syngas-to-light olefins process. (**A**) Selectivity of light olefins versus CO conversion data reported for FTTO (black squares) and OXZEO (blue triangles) processes (*6*–*14*, *16*, *18*–*24*, *30*), and this work (red circles). The red dashed line is a guide for the eye. (**B**) Scheme depicting the activity-selectivity trade-off caused by the entangled target (ketene intermediate-to-light olefins) and undesired secondary reactions (overhydrogenation to paraffins and oligomerization to C_{5+} hydrocarbons) over zeotypes.

also active for the secondary reactions of olefins, such as hydrogenation and oligomerization (*18*, *26*–*28*). Moreover, these secondary reactions are enhanced by the simultaneously strengthened acidity and increased active sites (*6*, *7*, *14*, *16*, *18*, *29*). For example, Li *et al.* showed that CO conversion was enhanced from 27 to 50% by increasing the Si content of SAPO-18, but the light-olefins selectivity declined from more than 80% to around 50% because of the simultaneous acceleration of secondary hydrogenation reactions (*18*). It remains a mechanistic challenge for conventional aluminosilicate and SAPO zeotypes.

Therefore, the goal is to terminate the reaction at the olefin formation step—that is, to disentangle the target reaction C–C coupling of ketene from the secondary reactions of hydrogenation and oligomerization of olefins (Fig. 1B). We report that Ge-substituted AlPO-18 is an effective strategy to achieve this goal and thus break the activity-selectivity trade-off. As a result, we obtained an unprecedented CO conversion of 85% simultaneously with 83% light-olefins selectivity among hydrocarbons. The light-olefins yield reaches as high as 48%, which nearly doubles the best values reported previously for FTTO and OXZEO processes.

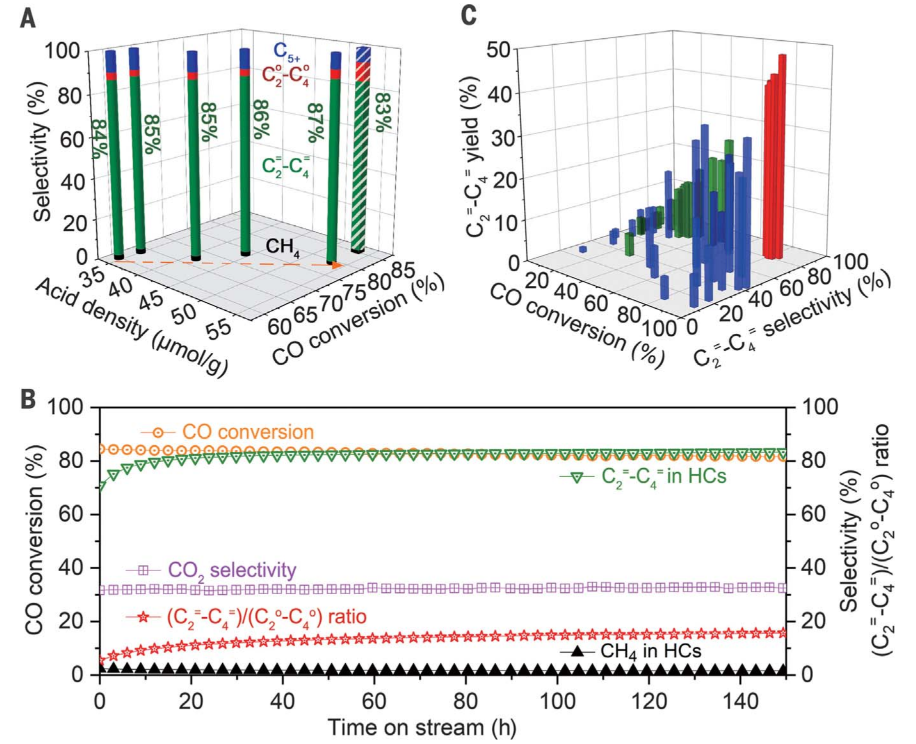


Fig. 2. Catalytic performance of $ZnCrO_x$–GeAPO-18 catalysts in syngas conversion. (**A**) Performance as a function of the medium-strength Brønsted acid site density of GeAPO-18. The dashed arrow line is a guide for the eye. Reaction condition: $ZnCrO_x$/GeAPO-18 = 2 (mass ratio), H_2/CO = 2.5, 410° and 430°C (prism filled pattern), 6 MPa and 1500 ml/(g_{cat}·hour). (**B**) Stability test of $ZnCrO_x$–GeAPO-$18_{0.027}$ at 430°C. (**C**) Performance of $ZnCrO_x$–GeAPO-18 (red columns) in comparison with previously reported FTTO (blue) and OXZEO catalysts (green) (*6*–*14*, *16*, *18*–*24*, *30*).

Germanium incorporation in AlPO-18

We introduced Ge into AlPO-18 by means of a hydrothermal method, with the Ge/Al ratio varying between 0.012 and 0.027 (denoted as the subscript "a"). The resulting GeAPO-18_a zeotypes exhibited a typical AEI crystal structure with no impurity phase and homogenous distribution of Ge species throughout the crystals (figs. S1 to S4 and table S2). Taking GeAPO-$18_{0.025}$, for example, x-ray absorption spectroscopy (XAS) of the Ge K-edge (fig. S5 and table S3) indicated that Ge species were dispersed as single sites being fourfold coordinated with O atoms within the AlPO-18 framework. Infrared analysis of the hydroxyl (OH) stretching vibration (fig. S6) and NH_3 adsorbed on GeAPO-18_a (fig. S7) confirmed the generation of Ge–OH–Al Brønsted acid sites created by the framework substitution of P by Ge atoms (figs. S8 and S9). A hydrothermal treatment experiment at 450°C demonstrated a good stability (fig. S10). NH_3 temperature-programmed desorption (TPD) indicated that the acid site density increased with the Ge/Al ratio up to 0.027, but the acid strength remained about the same (fig. S11 and table S4).

The GeAPO-18_a variants were physically mixed with $ZnCrO_x$ to form OXZEO composite catalysts for syngas conversion. CO conversion reached 60% upon introduction of a small amount of Ge (Ge/Al = 0.012), and light-olefins selectivity among hydrocarbons was 84% at 410°C. CO conversion rose almost linearly to 79%, with the density of the medium-strength acid sites increasing to 53 μmol/g, and light-olefins selectivity increased to 87% (Fig. 2A). CO conversion further increased to 85% at 430°C, and light-olefins selectivity declined only slightly to 83% (Fig. 2A and fig. S12). By contrast, AlPO-18 containing no Ge exhibited a low CO conversion of 24% and light-olefins selectivity of 63% under the same conditions (fig. S13 and table S5).

A test of 150 hours on stream showed that the catalyst delivered stable performance during this time frame (Fig. 2B, fig. S14, and table S6). The ratio of light-olefins to light-paraffins, denoted as $(C_2^=–C_4^=)/(C_2^o–C_4^o)$, remained at 16. The yield of light olefins reached as high as 48% (Fig. 2C), surpassing the previously reported best values for both FTTO and OXZEO catalysts by nearly a factor of 2 (*6*–*14*, *16*, *18*–*24*, *30*). The above results indicated that the "performance ceiling" shown in Fig. 1A is successfully broken, with CO conversion and light-olefins selectivity both exceeding 80%.

Catalyst comparison

To further illustrate the versatility of GeAPO-18 zeotypes, we compared them with corresponding catalysts substituted with silicon (SAPO-18) and with magnesium (MgAPO-18). Those two zeotypes were prepared with similar methods as that used for GeAPO-18 (supplementary materials) with a Si/Al ratio ranging from 0.011 to 0.029 and Mg/P ratio from 0.019 to 0.029, respectively. The resulting $ZnCrO_x$–MeAPO-18_a combinations were all tested in a wide temperature range of 410° to 480°C and space velocity between 1500 to 6000 ml/(g_{cat}·hour). The CO_2 selectivity appeared to be controlled thermodynamically (figs. S15 and S16). The red-orange-yellow region of the contour plots in Fig. 3, A to C, indicates light-olefins selectivity >80%, and black regions indicate selectivity <60%. The maps indicate that GeAPO-18 allowed the reaction to be operated over a wider temperature range, in which light-olefins selectivity and CO conversion simultaneously remained >80% with the $(C_2^=–C_4^=)/(C_2^o–C_4^o)$ ratio > 15 (Fig. 3A and fig. S17). A wider operation temperature range is highly desirable for applications. By contrast, the red-orange-yellow operation zone was much smaller with the classical SAPO-18 zeotypes (Fig. 3B) and almost nonexistent with MgAPO-18 (Fig. 3C). All SAPO-18_a, MgAPO-18_a, and GeAPO-18_a samples examined under different conditions in this work are detailed in Fig. 3D. It reveals that the activity-selectivity trade-off boundary cannot be exceeded with SAPO-18 and MgAPO-18

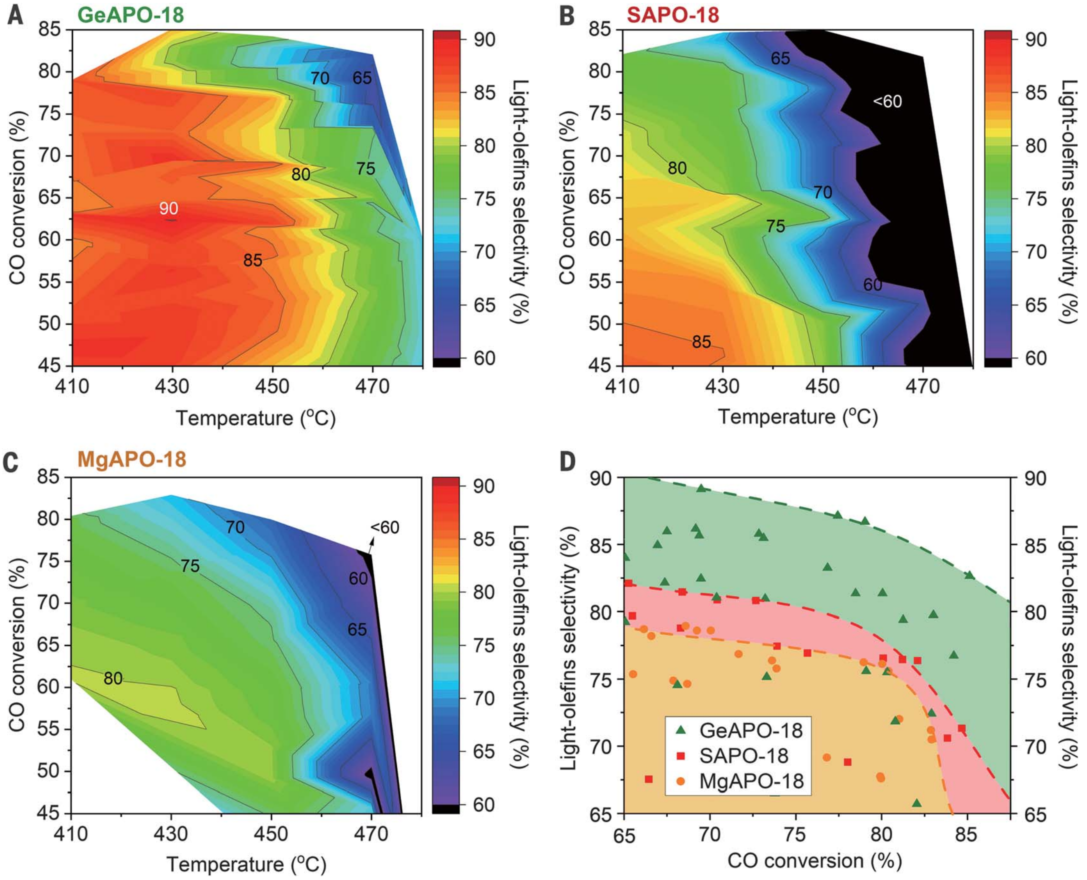


Fig. 3. Comparison of $ZnCrO_x$-MeAPO-18 in syngas conversion. Contour plots of catalytic performance. (**A**) GeAPO-18 with Ge/Al = 0.012 to 0.027. (**B**) SAPO-18 with Si/Al = 0.011 to 0.029. (**C**) MgAPO-18 with Mg/P = 0.019 to 0.029. (**D**) Selectivity of light olefins versus CO conversion, with the dashed lines being guides for the eye. Reaction condition: H_2/CO = 2.5, 410° to 480°C, 6 MPa, 1500 to 6000 ml/(g·hour).

by changing their acidity properties (varying Si/Al and Mg/P ratios) or the reaction conditions, but it can be breached by GeAPO-18.

Mechanistic studies

We compared GeAPO-$18_{0.025}$, SAPO-$18_{0.016}$, and MgAPO-$18_{0.029}$ to understand the mechanism underlying the high yield of olefins over the $ZnCrO_x$–GeAPO-18 catalysts because all three catalysts have a similar density of the medium-strength acid sites (table S4) as well as comparable diffusion efficiency (fig. S18). The NH_3-TPD profiles in fig. S11 show that the desorption corresponding to the medium-strength acid sites peaks at 235°C for GeAPO-$18_{0.025}$, which is 50°C lower than that of SAPO-$18_{0.016}$. In comparison, MgAPO-18 shows a broad high-temperature desorption feature (between 300° and 413°C). The desorption temperatures reveal that the strength of the acidity decreased in the sequence of MgAPO-18 > SAPO-18 > GeAPO-18. These results were consistent with the experimentally measured differential adsorption heat of NH_3 (table S7) and NH_3 binding energy on different Brønsted acid sites calculated with density functional theory (DFT) (table S8 and figs. S8 and S19), which were, respectively, 143 kJ/mol and –1.24 eV for GeAPO-$18_{0.025}$, 156 kJ/mol, and –1.40 eV for SAPO-$18_{0.016}$, and 193 kJ/mol and –1.71 eV for MgAPO-$18_{0.029}$.

Because the light-olefins yield improved with decreasing strength of the zeotype acidity, we examined the effect of the acidity on the adsorption and activation of H_2 and olefins by using ethylene as a probe molecule (table S9 and figs. S8 and S20). The calculated mean ethylene adsorption energy was 0.56 ± 0.05 eV on the Ge–OH–Al Brønsted acid sites (Fig. 4A), which was noticeably weaker than that on Si–OH–Al and Mg–OH–P sites. This order correlated well with the NH_3 binding energy trend, even considering the variation of the energies on different T-sites of each zeotype (Fig. 4A), and was further confirmed with the ethylene-TPD experiments (Fig. 4B). The H-D exchange experiments revealed a weaker H_2 activation activity on GeAPO-$18_{0.025}$ than that on SAPO-$18_{0.016}$ and MgAPO-$18_{0.029}$ (Fig. 4C) (*31*).

Both the weaker ethylene adsorption and H_2 activation over GeAPO-18 retard the secondary reactions of ethylene hydrogenation to ethane and its oligomerization to higher hydrocarbons C_{3+} (*18*). This is further demonstrated by ethylene conversion in the presence of hydrogen between 400° and 460°C over MeAPO-18 (Fig. 4D and fig. S21). GeAPO-18 exhibited lower activity for ethylene hydrogenation and oligomerization than those of the other two zeotypes. Although the presence of CO could significantly suppress the secondary reactions of olefins, GeAPO-18 still exhibited the weakest activity among the three zeotypes (fig. S22). The above results demonstrated that the Ge–OH–Al site can modulate the relative kinetics of C–C coupling of ketene to olefins and the subsequent over-hydrogenation and oligomerization of olefins and hence results in high selectivity for light olefins. The weaker oligomerization activity of Ge–OH–Al sites, as another advantage, led to a higher selectivity to more valuable ethylene and propylene compared with that of the optimized SAPO-18 under the same conditions (tables S10 and S11), which is of economic importance for industrial application. Therefore, CO conversion is improved by increasing the density of Ge–OH–Al sites and optimization of reaction conditions without degrading the light-olefins selectivity (Fig. 2A and fig. S23). An even higher activity can be expected if more Ge–OH–Al Brønsted acid sites could be generated as soon as a suitable preparative technique for enhanced heterometal substitution into zeotype frameworks becomes available (fig. S23).

Discussion

This study demonstrates that the key to breaking the activity and selectivity trade-off in CO hydrogenation to olefins lies in the disentanglement of the catalytically active sites for

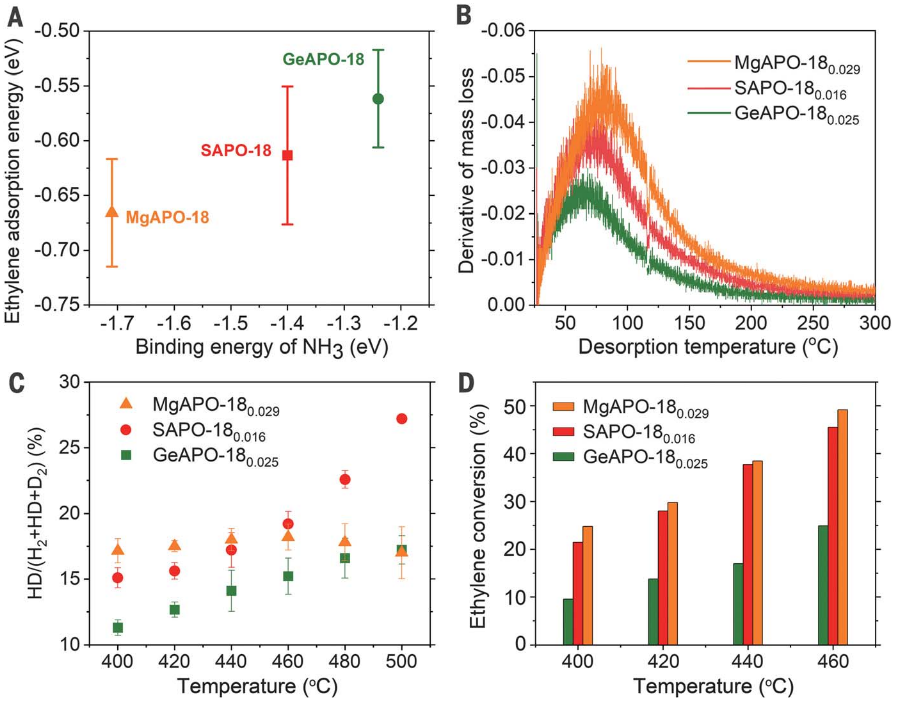


Fig. 4. Acidic strength of GeAPO-18 and its effects on the catalytic performance in comparison with SAPO-18 and MgAPO-18. (**A**) Calculated mean adsorption energy of ethylene versus the binding energy of NH_3. The error bars indicating the variation of adsorption energy between different T-sites of zeotypes (tables S8 and S9). (**B**) Ethylene-TPD measured by the derivative of the mass loss detected with intelligent gravimetric analysis. (**C**) H-D exchange activity represented by the relative $HD/(H_2+HD+D_2)$ concentration. (**D**) Ethylene hydrogenation at 400° to 460°C, 4 MPa, 200 ml/min, feed 1.2% C_2H_4-66.6% H_2-32.2% Ar.

the target and side reactions. It demonstrates that incorporation of Ge-substituted AlPO-18 provides an effective strategy within the OXZEO catalyst concept. Its weaker Ge–OH–Al Brønsted acid sites are largely inactive for the secondary reactions of hydrogenation and oligomerization of olefins, and increasing the density of Brønsted acid sites does not strengthen the acidity. Therefore, one can create more active sites to catalyze the conversion of ketene intermediates to light olefins and thus shift the thermodynamically limiting step of syngas conversion to ketene. As a result, 85% CO conversion and 83% light-olefins selectivity can be simultaneously obtained, resulting in an unprecedented light-olefins yield of 48% under optimized conditions. Furthermore, metal substitution constitutes a general strategy to mediate the strength and density of Brønsted acid sites in zeotype frameworks, which will enable the optimization of catalysts not only for CO but also CO_2 hydrogenation to value-added chemicals.

REFERENCES AND NOTES

1. U. Zavyalova, M. Holena, R. Schlögl, M. Baerns, *ChemCatChem* **3**, 1935–1947 (2011).
2. C. Zhan *et al*., *J. Am. Chem. Soc.* **142**, 14134–14141 (2020).
3. Q. Q. Guan *et al*., *Nat. Catal.* **4**, 840–849 (2021).
4. J. T. Hu *et al*., *Nat. Catal.* **4**, 242–250 (2021).
5. C. Wang *et al*., *J. Am. Chem. Soc.* **144**, 16855–16865 (2022).
6. F. Jiao *et al*., *Science* **351**, 1065–1068 (2016).
7. X. Pan, F. Jiao, D. Miao, X. Bao, *Chem. Rev.* **121**, 6588–6609 (2021).
8. H. M. Torres Galvis *et al*., *Science* **335**, 835–838 (2012).
9. Y. Xu *et al*., *Science* **371**, 610–613 (2021).
10. C. Wang *et al*., *Nat. Nanotechnol.* **17**, 714–720 (2022).
11. L. Zhong *et al*., *Nature* **538**, 84–87 (2016).
12. J. Xie *et al*., *Nat. Commun.* **10**, 167 (2019).
13. W. Fang *et al*., *Science* **377**, 406–410 (2022).
14. K. Cheng *et al*., *Angew. Chem. Int. Ed.* **55**, 4725–4728 (2016).
15. Y. Wang, *J. Energy Chem.* **25**, 169–170 (2016).
16. J. Su *et al*., *Nat. Commun.* **10**, 1297 (2019).
17. P. Gao, L. Zhong, B. Han, M. He, Y. Sun, *Angew. Chem. Int. Ed.* **61**, e202210095 (2022).
18. G. Li *et al*., *ACS Catal.* **10**, 12370–12375 (2020).
19. X. L. Liu *et al*., *ACS Catal.* **10**, 8303–8314 (2020).
20. L. Tan *et al*., *Chem. Sci.* **11**, 4097–4105 (2020).
21. N. Li *et al*., *Nat. Commun.* **13**, 2742 (2022).
22. H. Wang *et al*., *Natl. Sci. Rev.* **9**, nwac146 (2022).
23. P. Zhang *et al*., *Ind. Eng. Chem. Res.* **60**, 13214–13222 (2021).
24. Y. Ding *et al*., *ACS Catal.* **11**, 9729–9737 (2021).
25. Z. Z. Lai *et al*., *ACS Catal.* **11**, 12977–12988 (2021).
26. J. Kanai *et al*., *J. Catal.* **133**, 527–543 (1992).
27. S. Senger, L. Radom, *J. Am. Chem. Soc.* **122**, 2613–2620 (2000).
28. A. Vityuk *et al*., *ACS Catal.* **9**, 839–847 (2019).
29. X. Liu *et al*., *Chem. Sci.* **9**, 4708–4718 (2018).
30. H. M. Torres Galvis, K. P. de Jong, *ACS Catal.* **3**, 2130–2149 (2013).
31. H. H. Shin, S. McIntosh, *J. Mater. Chem. A* **1**, 7639–7647 (2013).

ACKNOWLEDGMENTS

We gratefully acknowledge Y. Li for his contribution in material synthesis. We thank beamline BL14W1 (Shanghai Synchrotron Radiation Facility) for the beam time and assistance with experiments. We thank X. Huang for discussion on XAS data analysis and H. Wang for discussion on Intelligent Gravimetric Analyzer data analysis. **Funding:** This work was supported by the Ministry of Science and Technology of China (2018YFA0704503), the National Natural Science Foundation of China (grants 91945302, 22272167, and 22002153), the Chinese Academy of Sciences (XDA21020400), the Youth Innovation Promotion Association of Chinese Academy of Sciences (2019184), and the Dalian Science and Technology Innovation Fund (2022RY02). **Author contributions:** F.J., G.L., C.X., Y.D., and J.L. performed the catalyst preparation. F.J., S.Q., and B.B. performed the catalytic reaction test. B.B., C.X., Z.J., W.L., T.P., and C.L. performed the characterization. Y.Y. and J.X. performed the DFT calculation. F.J., X.P., and X.B. designed the study, analyzed the data, and wrote the paper. **Competing interests:** A Chinese patent application (CN2018115750600) and six international patent applications (SA521421990; RU2021113024; ZA2021/02711; US17/286,903; IN202127019453; and EP19900221.3) through the Patent Cooperation Treaty (PCT) have been submitted by F.J., G.L., X.P., and X.B. **Data and materials availability:** All data are available in the manuscript and the supplementary materials. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adg2491
Materials and Methods
Supplementary Text
Figs. S1 to S23
Tables S1 to S12
References

Submitted 14 December 2022; accepted 17 April 2023
10.1126/science.adg2491

BIOENGINEERING

Restoration of natural thermal sensation in upper-limb amputees

Francesco Iberite[1], Jonathan Muheim[2], Outman Akouissi[2,3], Simon Gallo[4,5], Giulio Rognini[4,5], Federico Morosato[6], André Clerc[2], Magnus Kalff[2], Emanuele Gruppioni[6], Silvestro Micera[1,2]*†, Solaiman Shokur[1,2]*†

The use of hands for gathering rich sensory information is essential for proper interaction with the environment; therefore, the restoration of sensation is critical for reestablishing the sense of embodiment in hand amputees. Here, we show that a noninvasive wearable device can be used to provide thermal sensations on amputees' phantom hands. The device delivers thermal stimuli to specific regions of skin on their residual limb. These sensations were phenomenologically similar to those on the intact limbs and were stable over time. Using the device, the subjects could successfully exploit the thermal phantom hand maps to detect and discriminate different thermal stimuli. The use of a wearable device that provides thermal sensation can increase the sense of embodiment and improve life quality in hand amputees.

The human hand provides rich sensory information crucial for effective and safe manipulation (*1*, *2*). For prosthesis users, the ultimate goal is to restore sensation by creating a link with artificial sensors (*3*). Tactile sensations on users' phantom hands can be achieved by means of direct neural stimulation (*4*, *5*). A similar effect can also be noninvasively obtained by exploiting tactile phantom hand maps—regions of the residual limb able to elicit tactile projected sensations (*6*).

Although these approaches have been successfully used to restore haptic information, such as texture (*7*), shape (*8*), and stiffness (*9*), to the best of our knowledge, thermal feedback has not yet been restored. The very small size of thermal fibers does not allow reliable targeted direct electrical stimulation with state-of-the-art implantable neural interfaces (*10*). Here, we show that a noninvasive wearable device can be used to generate thermal sensations, similar to those experienced in an intact arm, in upper-limb amputees. To achieve this, we stimulated specific spots on the amputees' residuum and investigated whether the stimulation could elicit a thermal phantom hand map.

Twenty-seven adult (24 to 65 years old; four female) upper-limb amputee participants were included in the study (table S1). Twenty-two of the participants had transradial amputation, four of the participants had wrist disarticulations, and one participant had partial hand amputation. In all cases, amputations occurred at least 1 year prior. All but five participants used robotic prosthetic hands (two participants used body-powered prostheses, and three used cosmetic hands), and they all reported mild to no phantom limb pain on a visual analog scale (VAS < 3).

First, we investigated the presence of tactile phantom hand maps in all participants. Using the round side of a pen to stimulate different parts of the residual arm (Fig. 1A), we found evidence of projected sensation in 26 of the 27 participants (table S2).

[1]The BioRobotics Institute, Health Interdisciplinary Center, and Department of Excellence in Robotics and AI, Scuola Superiore Sant'Anna, 56127 Pisa, Italy. [2]Bertarelli Foundation Chair in Translational Neural Engineering, Neuro-X Institute, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland. [3]Bertarelli Foundation Chair in Neuroprosthetic Technology, Neuro-X Institute, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland. [4]Bertarelli Foundation Chair in Cognitive Neuroprosthetics, Neuro-X Institute, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland. [5]Metaphysiks Engineering SA, 1260 Nyon, Switzerland. [6]Centro Protesi INAIL, 40054 Vigorso di Budrio, Italy.
***Corresponding author. Email: silvestro.micera@epfl.ch (S.M.); solaiman.shokur@epfl.ch (S.S.)**
†These authors contributed equally to this work.

Presence of thermal phantom hand maps in upper-limb amputees

Next, we performed a thermal investigation on the residuum using a custom thermal device named MetaTouch (Metaphysiks, Lausanne) (Fig. 1B and supplementary materials). We found evidence of thermal phantom hand maps in 17 out of 27 participants. The sensation manifested in several points (up to four in some participants) with distinct somatotopic projection on the phantom hand and thermal description; for example, one participant reported "I feel you are cooling down my left index finger." For each spot on the residual arm, we tested three thermal stimuli, at 25°, 32°, and 37°C, and asked the participants to rate their thermal perception on a visual analog scale ranging from very cold to very hot (Fig. 1C).

The thermal perception was not uniform on all the tested spots; in some cases, the participants only perceived the hot stimuli, and in other cases, they only perceived the cold stimuli (fig. S1). We focused mainly on spots where the participants could perceive the cold, neutral, and warm stimuli (spots of coherent thermal perception) and where the projected sensations overlapped for the three stimuli (coherent locations). We considered these as thermal phantom spots that could be used for neuroprosthetics (Fig. 1D). We found evidence of thermal phantom spots in 15 of the 27 participants (table S2 and fig. S3), suggesting that in 56% (15/27) of the subjects, thermal phantom hand maps might be exploited for prosthetic use (Fig. 1E).

We also discovered that in some participants, the thermal phantom hand maps were present even in the absence of tactile stimulation. We used a chemical cooling agent (menthol cream) on one of the thermal phantom spots (*11*) of five participants. Among the participants, two reported thermal projected sensations (table S3). The cooling sensations were projected on similar regions of the phantom hand when we used the cooling agent and the MetaTouch (Fig. 1F).

Thermal phantom sensations are phenomenologically similar to those of the intact hand

To further investigate the phenomenological experience of the thermal projected sensations, we compared the sensation elicited by the MetaTouch on a thermal phantom spot with the physical stimulation on the mirrored intact body part with four amputee participants. Before the tests with the amputee participants, and to validate that the MetaTouch could deliver sensations similar to those produced by contact with real objects, we ran a control experiment with eight nonamputee subjects where the MetaTouch and the physical objects were placed on mirrored body parts. We chose the lower abdomen as the stimulation region; in a previous study (*12*), we reported that this is a region with particularly good sensitivity (better than that of the hand) to material detection. The MetaTouch produced the signature thermal drop of the skin in contact with one of the following three materials: copper, glass, and plastic (Fig. 2A). On the contralateral side, we used slabs constructed of the same three materials and of the same size as the MetaTouch thermode (3 cm by 4 cm). The participants had to report whether the two stimuli matched in terms of both location and thermal sensation. On average, the nonamputee participants reported that the physical and MetaTouch-mediated stimuli matched at 76.4%, with the highest rates of matching for plastic (86.8%), followed by glass (72.9%) and copper (68.75%) (Fig. 2B). In all cases, when we used nonmatching parameters (for example, physical copper on one side and MetaTouch-mediated plastic stimuli on the other), the probability of participants reporting that two stimuli matched was significantly lower (30.1%, $P = 0.004$, on average; 15%, $P = 0.004$, for plastic; 41.1%, $P = 0.008$, for glass; and 34.1%, $P = 0.039$, for copper; Friedman test).

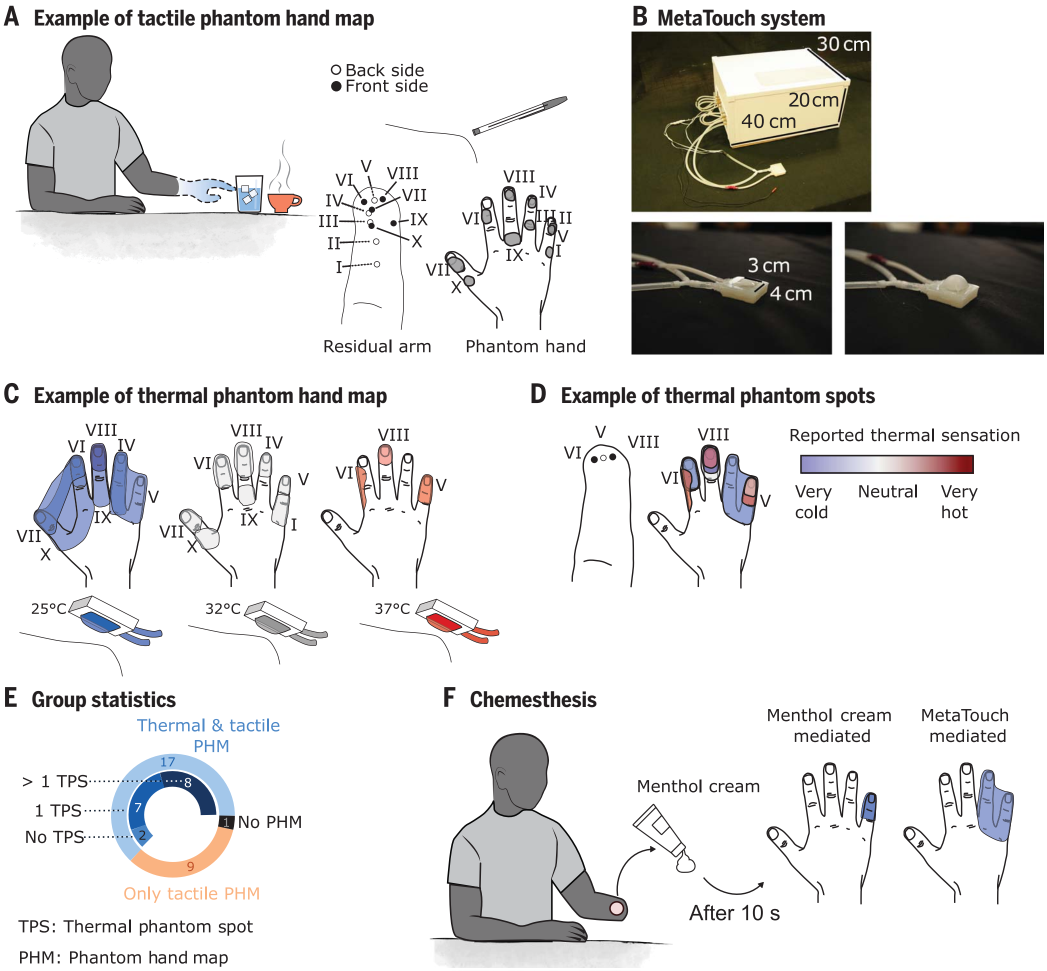


Fig. 1. Tactile and thermal phantom hand maps. (**A**) Tactile stimulation was performed on the residual arm using the round part of a pen. (Left) Top view of the residual arm, with spots on the front (black circles) and back (hashed circles) of the arm labeled, and (right) the corresponding tactile projection onto the phantom hand. (**B**) The MetaTouch system (Metaphysiks, Lausanne): a custom-made, hydraulic, thermal device that allows precise thermal display on a 3 cm by 4 cm inflatable silicone cell. (**C**) Reported projection on the phantom hand for stimulations performed on the residual arm [same points as in (A)] with the thermode set at 25°, 32°, or 37°C. The color corresponds to the participants' response, using a thermal visual analog scale, that ranges from very cold (dark blue) to very hot (dark red). (**D**) Example of three thermal phantom spots. (**E**) Group statistics for the 27 participants. (**F**) Comparison between a MetaTouch-mediated thermal sensation in the phantom hand and the sensation mediated by applying menthol cream. All examples in this figure are derived from participant P12.

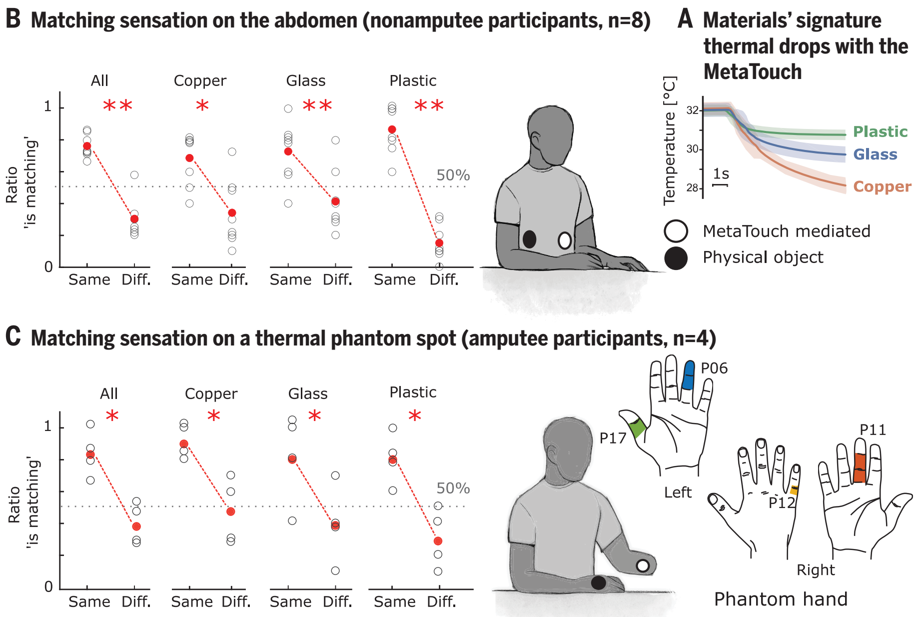


Fig. 2. Phenomenological experience. (**A**) MetaTouch-mediated thermal curves. (**B**) Healthy participants (n = 8) reported whether stimuli with a physical slab constructed of copper, glass, or plastic placed on their abdomen matched or did not match (in terms of the position and thermal properties) the MetaTouch-mediated stimuli on the mirrored side. (**C**) The test with amputee participants (n = 4) was performed with the MetaTouch placed on a thermal phantom spot on the residuum and physical objects placed on the mirrored intact body part. The set of physical and MetaTouch-mediated stimuli were not described to the participants. A gray dot represents, for each participant, the percentage of trials where they perceived the stimuli as matching, considering the trials where the stimuli were the same ("Same") and those that were different ("Diff."). The red dots denote averages. *P < 0.05, **P < 0.01; Friedman test.

Next, we ran the matching task with four amputees, all of whom had at least one thermal phantom spot (Fig. 2C). On average, the amputee participants reported that the physical and MetaTouch-mediated stimuli matched at 83.3%, a rate similar to what we had observed in nonamputee participants (76.4%). The highest rates in this case were found for copper (90%), followed by plastic and glass (both at 80%) (Fig. 2B). When we presented non-matching parameters, the probability of participants reporting that two stimuli matched

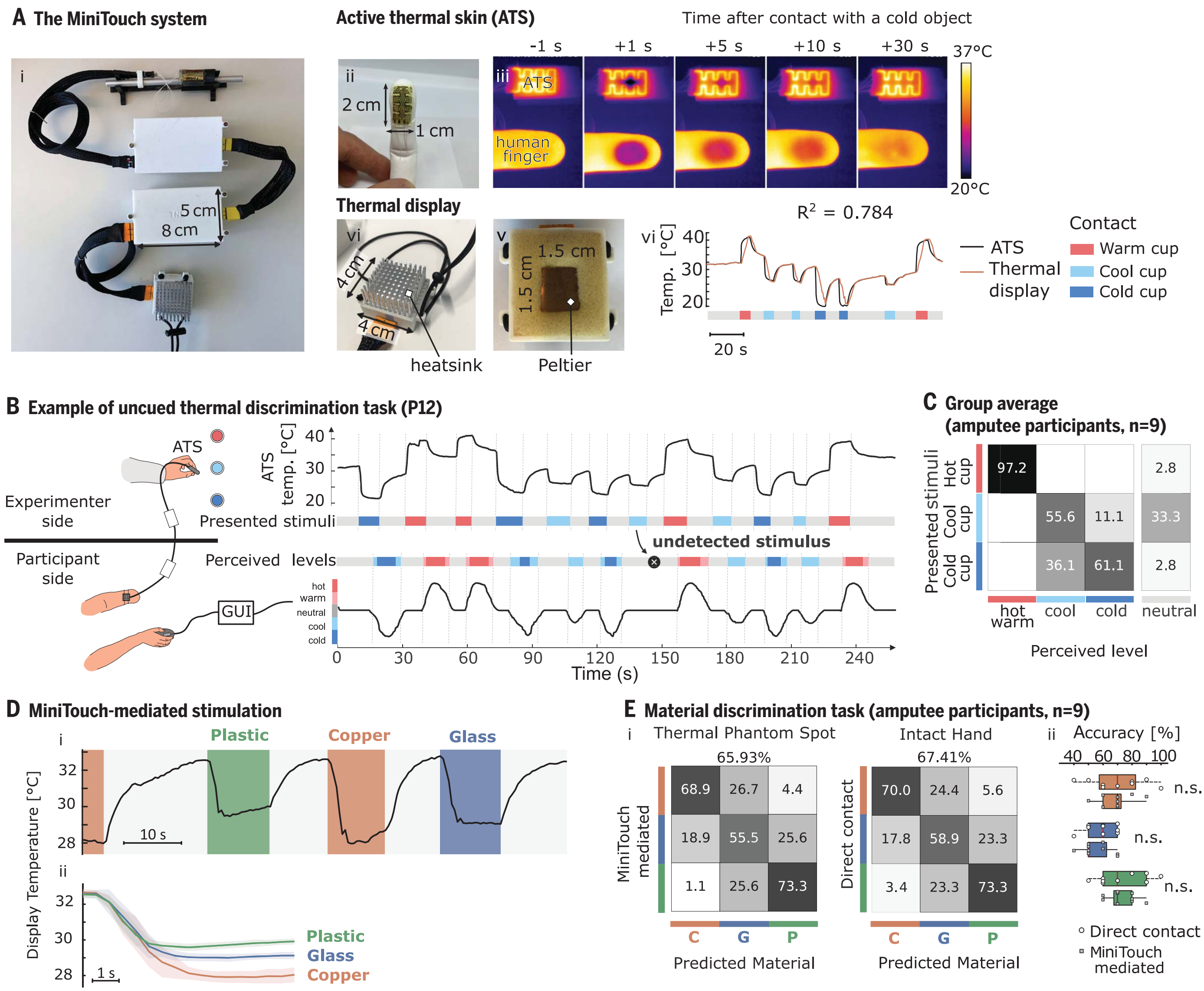


Fig. 3. Thermal feedback in sensory perception tasks. (**A**) Composition of the MiniTouch: (i) The controller (white boxes), sensor, and Peltier elements that stay in contact with the skin. (ii) The active thermal skin (ATS) sensor mimics temperature dynamics of the (iii) skin temperature. The heatsink element (iv) mounted on the Peltier elements (v) allows heat dissipation. (vi) The Peltier elements follow the temperature given by the ATS sensor. Red, light-blue, and dark-blue regions correspond to contact periods with hot, cool, and cold objects, respectively. Between contacts (gray regions), the ATS recovers to the skin baseline temperature (32°C). (**B**) The experimenter randomly placed the ATS in contact with objects at 15°, 24°, or 40°C, and participants were asked to report their sensation on a visual analog scale changing from cold to hot on a graphical user interface (GUI). (**C**) Confusion matrix. (**D**) (i) Temperature at the surface of the Peltier elements at the contact onset and offset with plastic, glass, and copper. (ii) Mean ± SD for all three materials (10 trials per material). (**E**) (i) Confusion matrix for three-class forced-choice material identification for MiniTouch–mediated materials and physical objects (mean score over nine participants) and (ii) their comparison per material type. n.s., nonsignificant; Wilcoxon rank sum test.

was again significantly lower (39.17%, $P = 0.04$, on average; 47.5%, $P = 0.04$, for copper; 30.0%, $P = 0.04$, for plastic; and 40.0%, $P = 0.04$, for glass; Friedman test). Our results show that thermal phantom sensations are phenomenologically similar to those of the intact hand.

Delivering thermal phantom sensations through a wearable device

We then investigated whether our approach could be used to deliver stable sensations over time. We designed a custom-made wearable device, the MiniTouch system (Fig. 3A and supplementary materials). The MiniTouch uses two Peltier elements to cover a 15 mm by 15 mm stimulation area. An important feature of the MiniTouch is the active thermal skin (ATS) sensor, which is a flexible thin-film (200-μm) device with heating and sensing properties designed to be mounted on existing prostheses. The ATS sensor keeps the temperature at sensor's surface at 32°C under a constant power supply, mimicking the thermal response of the skin.

We tested the MiniTouch in nine amputees for whom we had previously identified the presence of thermal phantom spots; two different thermal discrimination tasks were used. In the first test, we placed the MiniTouch on one of the thermal phantom spots and asked the participants to continuously report their thermal sensation using a slider on a graphical user interface (movie S1). The experimenter

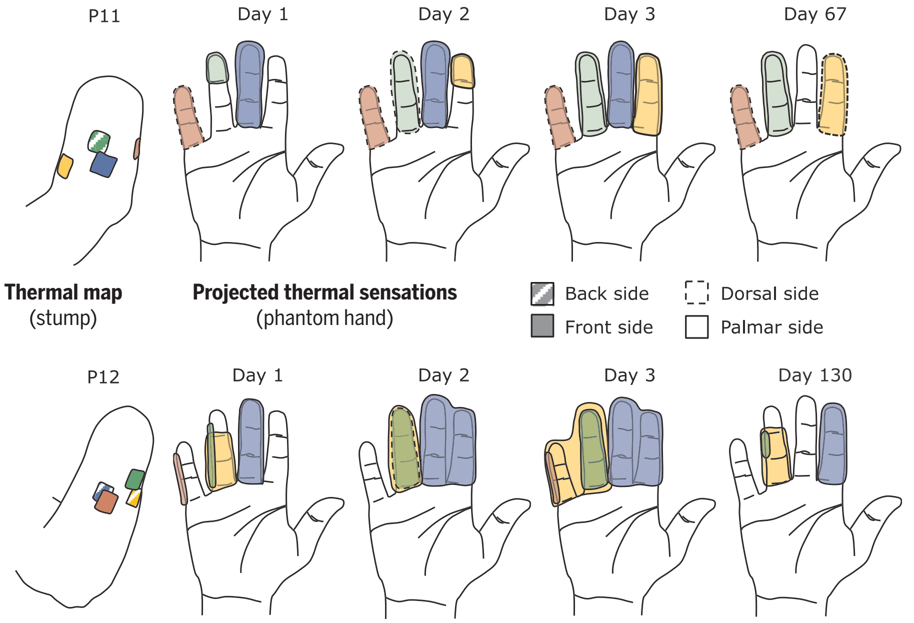


Fig. 4. Stability test. Examples of spots on the residual arm with stable projection over 3 days and after 67 and 130 days for P11 and P12, respectively. The temperature of the thermode (MiniTouch) was set to 25°C. On the residual arm, dashed regions correspond to the back side; on the phantom hand, solid lines represent projections on the palmar side (dashed lines indicate projections on the dorsal side).

randomly placed the ATS sensor in contact with one of three glasses of water at different temperatures (15°, 24°, or 40°C). Overall, 24 uncued events were presented for each participant (four repetitions for each glass and their recovery to baseline) (Fig. 3B).

On average, the participants detected 97.2% of the warm/hot (40°C) and cold (15°C) stimuli; for the cold stimuli, 61.1% were reported as cold and 36.1% as cool. For the cool stimuli (24°C), 66.7% were detected (55.6% were reported as cool and 11.1% as cold) (Fig. 3C). Considering a conservative estimate of four possible classes (cold, cool, neutral, and warm/hot) and chance level with $P < 0.01$ at 35.2% (*13*), we found that the participants were significantly above chance in discriminating between the stimuli (group mean accuracy: 71.3%). Participants detected, on average, 83.3% of the recovery-to-baseline events. Individual score analysis (fig. S4) shows that, in some cases, participants' perception was attenuated [for example, one participant (P19) rated three out of four times the cool stimuli as neutral and the cold stimuli as cool], suggesting that a simple gain factor between the ATS and thermode might have resulted in even better scores.

In the second test, the experimenter randomly placed the ATS sensor on slabs of copper, glass, or plastic and asked the subjects to identify the material (Fig. 3D and movie S2). Considering that the chance level for a three-label classification, with statistical significance of the identification with $P < 0.01$, was 40.0% (*13*), we again found that participants' detection rates were, on average, significantly above chance (65.93%, Fig. 3E, and see fig. S5 for individual scores). In addition, detection accuracy using the MiniTouch was similar to accuracy measured in the same participant touching physical objects with the intact hand (67.41%), and no significant difference could be found between (i) detection with MiniTouch on the phantom hand and (ii) physical objects on the intact hand for any of the materials: copper [mean score for (i): 68.9%; mean score for (ii): 70%; $P = 0.92$, Wilcoxon rank sum test], glass (55.6% versus 58.9%, $P = 0.50$), and plastic (73.3% versus 73.3%, $P = 0.92$).

Thermal phantom sensations are partially stable over months

As the last step, we tested the MiniTouch over a 3-day experiment in two participants, P11 and P12, for whom we had found thermal phantom spots. We placed a two-dimensional grid on the residual arm using a temporary tattoo (fig. S2). We used the grid to perform a systematic investigation of the tactile and thermal phantom hand maps. The phantom hand maps showed an indication of partial stability over consecutive days. For participant P11, we considered 37 locations on the residual arm, of which 11 were stable (their projection onto the phantom hand overlapped); for P12, across 12 locations, three remained stable. The probability that the stable points were given by chance was 7.7×10^{-7} for P11 and 0.003 for P12 (bootstrap method, supplementary materials). Representative examples of these spots for both subjects are given in Fig. 4. A follow-up test after 2 months for P11 and after 4 months for P12 confirmed evidence of partial stability. The stability of the projected points on the residual arm was also confirmed by a simplified yes/no questionnaire (supplementary materials).

Outlook

Altogether, our results show that thermal phantom hand maps can be applied in a partially stable and reliable way to restore thermal sensation in upper-limb amputees. We developed a relatively easy-to-implement solution for thermal restoration through a noninvasive approach that provided intuitive and phenomenologically close-to-natural sensation. Our approach might ultimately improve prosthetic users' experience, creating a foundation for affective touch (*14*, *15*), perception of wetness (*16*), and increased embodiment (*17*). Our approach could also improve sensation naturalness when integrated with existing solutions for tactile restoration by means of nerve stimulation (*4*, *5*, *8*) and provide a multimodal and more effective sensory experience to the users.

REFERENCES AND NOTES

1. A. R. Sobinov, S. J. Bensmaia, *Nat. Rev. Neurosci.* **22**, 741–757 (2021).
2. R. S. Johansson, J. R. Flanagan, *Nat. Rev. Neurosci.* **10**, 345–359 (2009).
3. S. J. Bensmaia, D. J. Tyler, S. Micera, *Nat. Biomed. Eng.* **2020**, 1–13 (2020).
4. M. Ortiz-Catalan, E. Mastinu, P. Sassu, O. Aszmann, R. Brånemark, *N. Engl. J. Med.* **382**, 1732–1738 (2020).
5. G. Valle *et al*., *Neuron* **100**, 37–45.e7 (2018).
6. B. Stephens-Fripp, G. Alici, R. Mutlu, *IEEE Access* **6**, 6878–6899 (2018).
7. C. M. Oddo *et al*., *eLife* **5**, e09148 (2016).
8. S. Raspopovic *et al*., *Sci. Transl. Med.* **6**, 222ra19 (2014).
9. D. W. Tan *et al*., *Sci. Transl. Med.* **6**, 257ra138 (2014).
10. S. Raspopovic, G. Valle, F. M. Petrini, *Nat. Mater.* **20**, 925–939 (2021).
11. D. M. Bautista *et al*., *Nature* **448**, 204–208 (2007).
12. M. N. Kalff, S. Shokur, E. F. Lavado, S. Micera, "Material surface detection on various body parts: A preliminary study for temperature substitution for upper arm amputees," International IEEE/EMBS Conference on Neural Engineering (NER), 4 to 6 May 2021, pp. 195–198.
13. E. Combrisson, K. Jerbi, *J. Neurosci. Methods* **250**, 126–136 (2015).
14. R. Ackerley *et al*., *J. Neurosci.* **34**, 2879–2883 (2014).
15. L. S. Löken, J. Wessberg, I. Morrison, F. McGlone, H. Olausson, *Nat. Neurosci.* **12**, 547–548 (2009).
16. D. Filingeri, D. Fournet, S. Hodder, G. Havenith, *J. Neurophysiol.* **112**, 1457–1469 (2014).
17. M. P. M. Kammers, K. Rose, P. Haggard, *Neuropsychologia* **49**, 1316–1321 (2011).

ACKNOWLEDGMENTS

We thank A. d'Alicandro, I. Furfaro, A. Sorriento, and E. Cigna for their valuable help, and we thank the subjects who freely donated their time to the study. We thank M. Herbelin for help with the illustrations. **Funding:** This work was supported by the Bertarelli Foundation (including the Catalyst programme); the Swiss National Science Foundation through the National Centre of Competence in Research (NCCR) Robotics and the CHRONOS project; and the European Union's Horizon 2020 research and innovation program under Marie Skłodowska-Curie grant

754354. This work was also supported by the Horizon Europe Research & Innovation Programme under grant 101092612 (Social and hUman ceNtered XR - SUN project) the #NEXTGENERATIONEU (NGEU) and partially funded by the Ministry of University and Research (MUR), National Recovery and Resilience Plan (NRRP) with two projects: MNESYS (PE0000006)–A Multiscale integrated approach to the study of the nervous system in health and disease (DN. 1553 11.10.2022) and THE (IECS00000017)–Tuscany Health Ecosystem (DN. 1553 11.10.2022). **Author contributions:** Conceptualization: F.I., J.M., E.G., S.M., and S.S. Data collection: F.I., J.M., F.M., M.K., and S.S. Data analysis: F.I., J.M., S.M., and S.S. Development of MetaTouch: S.G. and G.R. Characterization of MetaTouch: F.I. and M.K. Development of MiniTouch: J.M., O.A., A.C., and S.S. Supervision: S.M. and S.S. Writing – original draft: F.I., J.M., S.M., and S.S. Writing – review & editing: All authors. **Competing interests:** S.M. holds shares in SensArs, which aims to develop bionic limbs for amputees. G.R. and S.G. hold shares in Metaphysiks, which builds thermal devices. F.I., J.M., O.A., A.C., S.M., and S.S. are coinventors of a thermal sensing device and sensory feedback system and method using said thermal sensing device (application number EP22207038.5). **Data and materials availability:** All data are available in the main text or the supplementary materials. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adf6121
Materials and Methods
Figs. S1 to S6
Tables S1 to S3
References
MDAR Reproducibility Checklist
Movies S1 and S2

View/request a protocol for this paper from *Bio-protocol*.

Submitted 2 November 2022; resubmitted 22 February 2023
Accepted 5 April 2023
10.1126/science.adf6121

BIOMATERIALS

Neuromorphic sensorimotor loop embodied by monolithically integrated, low-voltage, soft e-skin

Weichen Wang[1]†, Yuanwen Jiang[2]†, Donglai Zhong[2]†, Zhitao Zhang[2]†, Snehashis Choudhury[2], Jian-Cheng Lai[2], Huaxin Gong[2], Simiao Niu[2], Xuzhou Yan[2], Yu Zheng[3], Chien-Chung Shih[2], Rui Ning[1], Qing Lin[4], Deling Li[5,6], Yun-Hi Kim[7], Jingwan Kim[7], Yi-Xuan Wang[2], Chuanzhen Zhao[2], Chengyi Xu[2], Xiaozhou Ji[2], Yuya Nishio[4], Hao Lyu[2], Jeffrey B.-H. Tok[2], Zhenan Bao[2]*

Artificial skin that simultaneously mimics sensory feedback and mechanical properties of natural skin holds substantial promise for next-generation robotic and medical devices. However, achieving such a biomimetic system that can seamlessly integrate with the human body remains a challenge. Through rational design and engineering of material properties, device structures, and system architectures, we realized a monolithic soft prosthetic electronic skin (e-skin). It is capable of multimodal perception, neuromorphic pulse-train signal generation, and closed-loop actuation. With a trilayer, high-permittivity elastomeric dielectric, we achieved a low subthreshold swing comparable to that of polycrystalline silicon transistors, a low operation voltage, low power consumption, and medium-scale circuit integration complexity for stretchable organic devices. Our e-skin mimics the biological sensorimotor loop, whereby a solid-state synaptic transistor elicits stronger actuation when a stimulus of increasing pressure is applied.

Skin plays a vital role as the primary interface and layer of protection between the human body and the surrounding world (*1*). The combination of sensory capabilities and mechanical softness of skin not only enables effortless perceptions and responses to various external stimuli but also allows us to perform complex tasks in both dynamic and unstructured environments (Fig. 1A) (*2*). Unfortunately, patients suffering from skin damage or amputation may have severely perturbed perception-action loops, causing them to struggle even with simple tasks such as object grasping. Although prosthetic limbs can partially restore motor functions (*3*), their limitations—such as phantom limb pain, uncontrolled movements, and poor dexterity—need to be addressed through the incorporation of sensory feedback and tissue compliance (*4*–*6*). To enable a natural human-machine interface, efforts have been made to confer (e.g., multimodal sensation, neuromorphic signal processing, closed-loop perception and actuation) (*5*, *7*–*11*) on electronic systems with skin-like properties (e.g., soft and deformable) (*12*–*15*).

Complex silicon (Si) circuits have been developed to mimic the peripheral nervous system and embody skin-like perceptual capability in prosthetics and robotics (*16*). However, building a monolithically integrated, soft electronic skin (e-skin) that combines all the desired electrical and mechanical features within one single device platform remains a challenge. Progress has been made in mimicking biological sensory functions to replace bulky and rigid systems with soft electronic materials for high-fidelity sensation (*9*, *17*, *18*). Despite the advancements in flexible electronics and circuit fabrication, their applications in e-skin have been limited to active-matrix arrays for reducing interconnect wirings and/or as signal amplification for sensors (*19*–*22*). An e-skin for neuroprosthetics needs to provide nerve-like pulse-train signal patterns to enable natural sensations and to modulate the corresponding motor activities (*2*, *5*). Even though piezoelectric, triboelectric, and ionic pressure sensors can generate spike train–like signal patterns according to the frequencies of applied forces (*23*), biological sensory receptors use frequency-modulated signals to encode the input amplitude information. This requires signal digitalization and conditioning circuits to convert analog signals from sensors to spike-train signal patterns and artificial synapses to regulate current amplitude to trigger body movements (Fig. 1B) (*10*, *11*, *16*, *24*).

With recent developments in soft electronic device fabrication, the feasibility of wiring rigid and soft components into hybrid e-skin systems has been demonstrated (*10*, *11*). However, it is desirable to construct a monolithic soft e-skin system because soft electronics have the advantages of good tissue conformability, minimal invasiveness, and imperceptibility (*9*, *17*). To meet the stringent requirements of a soft e-skin system with the distinctive advantages of biointegration (*12*), several fundamental challenges need to be overcome for existing soft electronics. On the materials level, they need to have tissue-like mechanical properties and excellent stability in the physiological environment (*25*, *26*). On the device level, safe and efficient on-body operation using a low operating voltage with excellent electronic characteristics, such as high charge-carrier mobility and small subthreshold swing (*27*), is needed. On the system level, signal conditioning circuits with sufficient transistor counts, logic depth, and functional complexity are required (*3*). Finally, on the application level, soft e-skin requires biomimetic encoding of sensory information and neuromorphic actuation at the device-bio interface for natural sensation and low-power consumption (*2*, *6*, *28*).

We report a monolithically integrated, soft, and low-voltage-driven e-skin system, without any rigid electronic components, that can emulate the sensory feedback functions of biological skin, including multimodal reception, nerve-like pulse-train signal conditioning, and closed-loop actuation (Fig. 1, C and D, and table S1). Specifically, a trilayer, high–permittivity (κ), stretchable dielectric design is required to reduce the driving voltage of stretchable circuits down to a few volts for safe and energy-efficient on-body operation.

Low-voltage-driven, high-performance, stretchable organic transistors

Existing intrinsically stretchable organic electronics still require a high operation voltage (30 to 100 V) (*29*, *30*), which generates safety and power consumption concerns. Reducing the driving voltage requires increasing the gate capacitance, which requires a thin dielectric layer that has a high-dielectric constant (Fig. 2Ai) (*31*–*33*). Among available dielectric elastomers, nitrile-butadiene rubber (NBR) is a promising candidate (fig. S1). Unlike ionic polymer dielectrics that suffer from a decrease in permittivity with increased frequency (*31*), the high permittivity of NBR (κ of ~28), which is induced by polarization of nitrile groups, can be well maintained within a wide frequency range up to 10^6 Hz (Fig. 2B) (*34*). This property is essential for the stable operation of transistors and circuits. We developed a direct photopatterning method down to 2 μm with an ultraviolet light–triggered azide–cross-linking reaction (fig. S2). The cross-linked NBR is solvent resistant, which is desirable for subsequent processing. However, the transistor transfer curve showed large hysteresis and low mobility (fig. S3). We attribute this to the energetic disorders at the semiconductor-dielectric interfaces that are induced by the highly polar nitrile groups. This phenomenon is known to occur for other organic transistors with high-κ dielectrics (*35*). Such a trade-off between a high κ value and high trap density makes it challenging to realize both low driving voltage and high carrier mobility in stretchable organic devices.

[1]Department of Materials Science and Engineering, Stanford University, Stanford, CA 94305, USA. [2]Department of Chemical Engineering, Stanford University, Stanford, CA 94305, USA. [3]Department of Chemistry, Stanford University, Stanford, CA 94305, USA. [4]Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA. [5]Department of Radiology, Molecular Imaging Program at Stanford (MIPS), Stanford University, Stanford, CA 94305, USA. [6]Department of Neurosurgery, Beijing Tiantan Hospital, Beijing Neurosurgical Institute, Capital Medical University, Beijing 100070, China. [7]Department of Chemistry and RINS, Gyeongsang National University, Jinju 660-701, South Korea.
***Corresponding author. Email: zbao@stanford.edu**
†These authors contributed equally to this work.

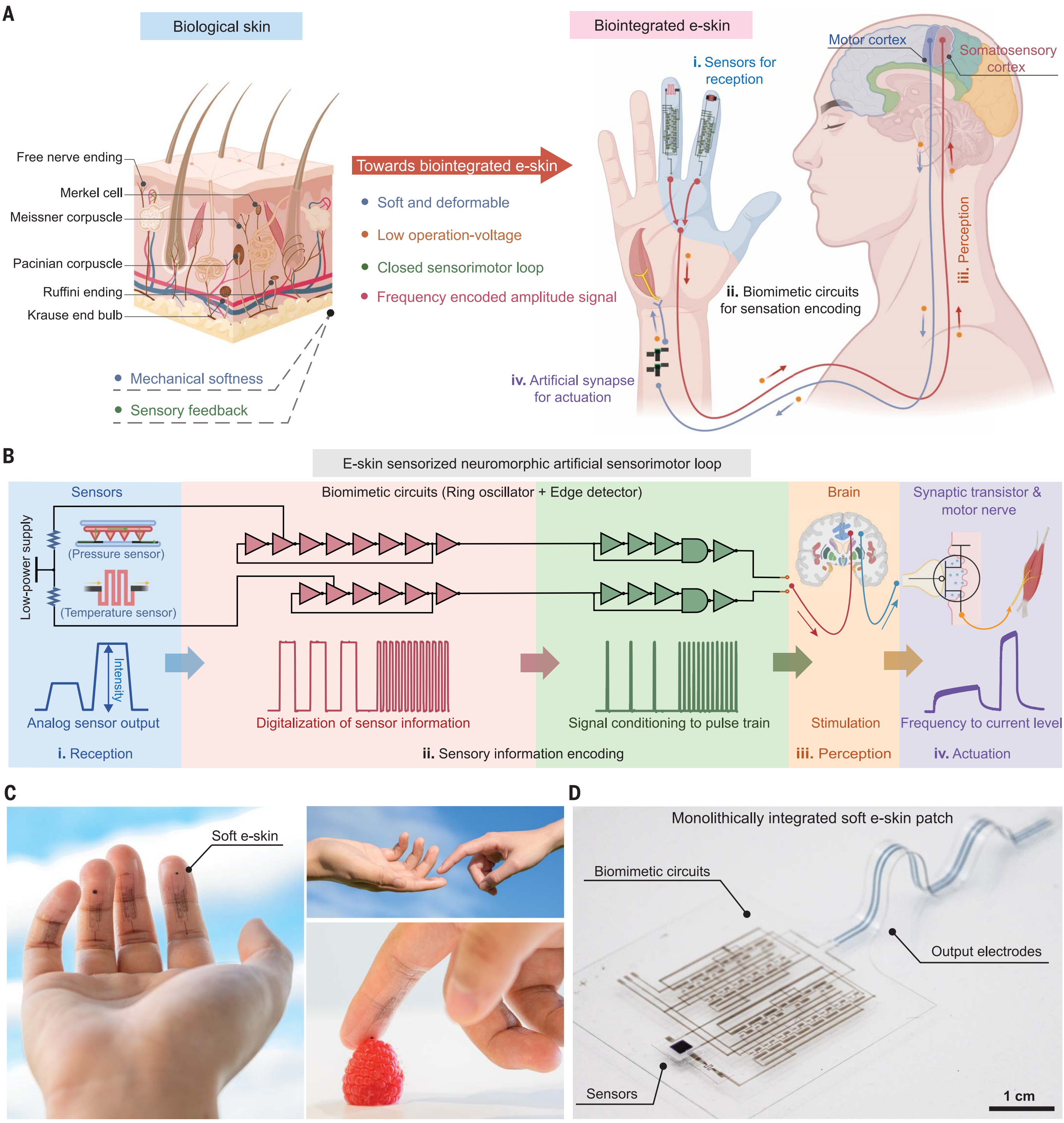


Fig. 1. Low-voltage-driven artificial soft e-skin system enables biomimetic bidirectional signal transmission. (**A**) On the left is an illustration showing the combination of mechanical softness and sensory feedback in biological skin. The locations of the following receptors in skin are labeled: free nerve ending (pain and temperature), Merkel cell (touch), Meissner corpuscle (sensitive touch), Pacinian corpuscle (vibration and pressure), Ruffini ending (touch and pressure), and Krause end bulb (touch). On the right is a schematic diagram illustrating the biointegrated e-skin system for an artificial sensorimotor loop. Artificial sensory receptors can respond to different external stimuli (temperature and pressure). Low-voltage-driven circuits consisting of ROs and EDs can encode sensor signals into pulse trains for the artificial sensory process. Stretchable artificial synapses can actuate downstream muscle movements. (**B**) Schematic showing the overall flow and e-skin components for the implementation of the artificial sensorimotor loop. (**C**) Photos showing the excellent skin conformability of a biointegrated e-skin system consisting of a temperature sensor, a pressure sensor, and two sets of RO-ED integrated circuits. Touching fingertips and a finger pressing on a raspberry illustrate the capability of sensation and mechanical softness of our developed e-skin system. (**D**) Photo showing the components of a monolithically integrated, soft e-skin patch. [Illustrations in (A) and (B) were created with BioRender.com; photos in (C) and (D) were taken by J.-C.L.]

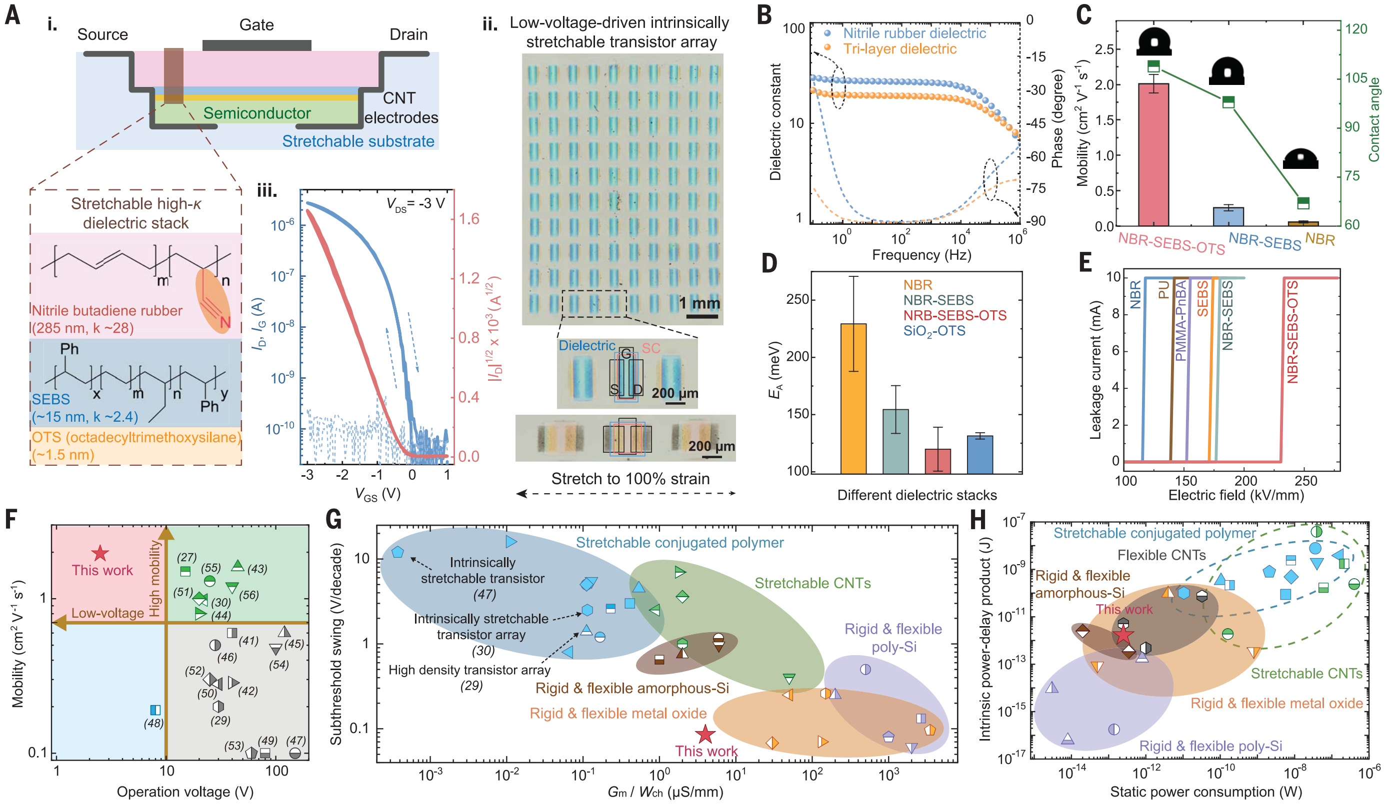


Fig. 2. High-κ, trilayer dielectrics for high performance and low-voltage stretchable organic transistors and circuits. (**A**) Panel (i) shows a schematic of the device structure and materials for high-κ, trilayer dielectric and low-voltage-driven transistors. x, m, n, and y represent the degree of polymerization. CNT, carbon nanotube; k, dielectric constant; Ph, phenyl. Panel (ii) is an optical microscope image showing the transistor array with a device density of 330 transistors per cm^2. Magnified images show the transistors under 0% (top) and 100% (bottom) strains. S, D, and G represent the source, drain, and gate electrodes, respectively. SC, semiconductor. Panel (iii) shows the transfer characteristics of a transistor from the batch with the highest mobility operating under 3 V. Typically, average mobility falls between ~0.7 and 1.5 cm^2 V^{-1} s^{-1}. A maximum average value of 2.01 cm^2 V^{-1} s^{-1} was obtained from a batch of transistor arrays containing 10 transistors each. I_D, drain current; I_G, gate current; V_{DS}, drain-source voltage; V_{GS}, gate-source voltage. (**B**) Dielectric constants and phase angles of a single-layer NBR and trilayer dielectrics under different frequencies. The dashed ovals with arrows indicate the dielectric phase angle. (**C**) Transistor mobility (for details, see fig. S3) and water contact angle for different dielectrics. The black graphics represent the testing results of water contact angle. (**D**) Comparison of the E_A for different dielectrics (averaged from three transistors from one batch). (**E**) Comparison of the dielectric strength and breakdown voltage of different kinds of elastomeric dielectrics with those of our trilayer dielectric. PMMA-PnBA, poly(methyl methacrylate)-*b*-poly(n-butyl acrylate); PU, polyurethane. (**F**) Comparison of previously reported stretchable polymeric transistors with our work in terms of operation voltage and mobility. (**G** and **H**) Comparison of previously reported transistors using conjugated polymer, carbon nanotube, metal oxide, amorphous Si, or poly-Si as the active layer with our low-voltage-driven transistor in terms of subthreshold swing and transconductance normalized by channel width (G_m/W_{ch}), and single-transistor-level intrinsic power-delay product and static power consumption (H). The supply voltage for the power consumption calculation is the voltage needed for an on-off current ratio of 10^4. The shaded and dashed circles group related references together. The color of the shade or circle line corresponds to the captions with the same color. All error bars refer to standard deviations. In (F) to (H), the different shapes and patterns of the symbols are used as a visual aide to distinguish different transistors from existing works (see details in figs. S24 and S26).

To overcome this challenge, we developed a trilayer dielectric by passivating the high-κ NBR with an ultrathin, nonpolar poly(styrene-ethylene-butylene-styrene) (SEBS) elastomer coating (~15 nm), followed by a hydrophobic octadecyltrimethoxysilane (OTS) molecular modification (see detailed thickness optimization in fig. S4) (Fig. 2A). This dielectric stack with a suitable surface energy induces a desirable nanoconfined morphology in the stretchable semiconductor layer, allowing for a high charge-carrier mobility (Fig. 2C and figs. S5 and S6) (*27*). In addition, it offers a smooth surface (average roughness <0.17 nm) with low trap density to decrease the charge-carrier scattering (figs. S7 to S9). Therefore, compared with the single-layer NBR dielectric devices, in the trilayer design, the carrier mobility was boosted by around 50-fold while the low driving voltage was maintained (Fig. 2C). To quantify the interfacial trap density (*35*), we measured the activation energy for charge transport (E_A) of various dielectric designs. The combination of NBR-SEBS-OTS with a directly spin-casted semiconductor shows the lowest E_A because of the optimal dielectric-semiconductor interface and semiconductor morphology (Fig. 2D). Another benefit afforded by our trilayer design is the substantially enhanced dielectric strength under both direct and alternating voltage biasing, thus allowing the use of a thinner dielectric layer to further increase the gate capacitance. After the deposition of the SEBS and OTS layers, we observed a steady increase in the breakdown voltage, indicating less pinholes (Fig. 2E and fig. S10). The minimum thickness with a high yield in transistor arrays (>80%) is ~300 nm (285 nm of NBR plus 15 nm of SEBS), whose capacitance value is 30-fold higher than the SEBS-only dielectric of 1.2 μm (*30*). Compared with other previously reported elastomeric dielectrics (*29*, *31*), our trilayer design allows low operation voltage and high carrier mobility at the same time (Fig. 2F and table S2).

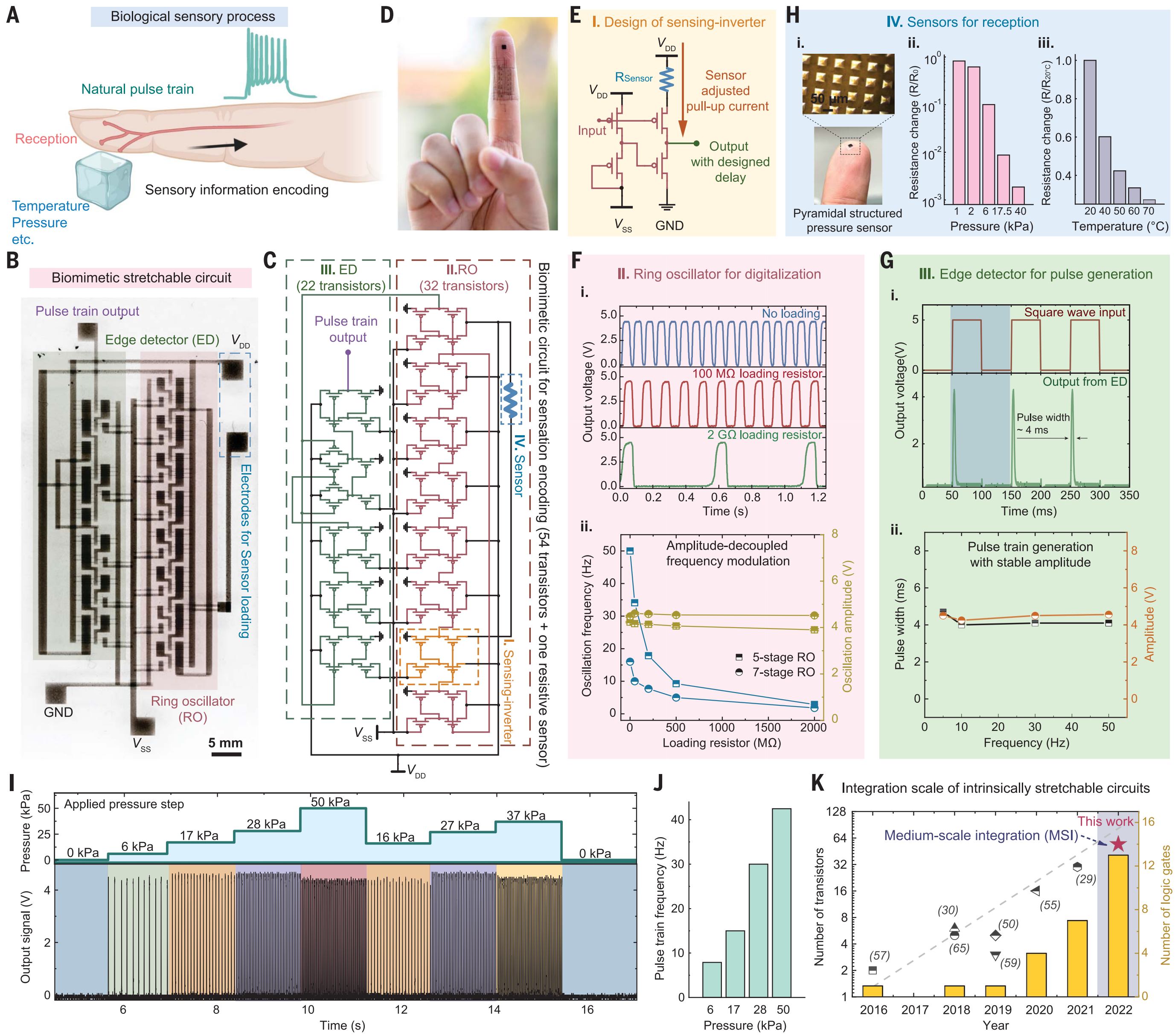


Fig. 3. Low-voltage-driven, soft circuit system for generating biomimetic pulse trains. (**A**) Schematic diagram showing the natural perception process. [Illustration was created with BioRender.com.] (**B**) Optical microscopic image of a stretchable seven-stage RO-ED circuit. GND, ground; V_{SS}, negative power voltage. (**C**) Transistor circuit diagram of the biomimetic sensor-circuit system for sensory information encoding. (**D**) Photo of a soft e-skin attached to a finger. (**E**) Circuit diagram and working mechanism of the sensing inverter. (**F**) Panel (i) presents transfer curves showing the amplitude-decoupled frequency modulation of a seven-stage RO with different loading resistor values. The RO output frequency changed from 16 Hz at 0 ohms to 1.8 Hz at 2 gigohms. Panel (ii) shows the oscillation frequencies and amplitudes for a five- and a seven-stage RO while loading different resistors. (**G**) Panel (i) presents input (top) and output (bottom) signals from an ED that produces a pulse width of ~4 ms. Panel (ii) shows the averaged pulse widths (squares) and amplitudes (circles) from three different individual pulses of an ED under different input signal frequencies. Inputs are as follows: square wave, amplitude 5 V, and 50% duty cycle. (**H**) Panel (i) contains a photo and magnification showing a pressure sensor with pyramidal structures. Resistance (R) changes of a pressure sensor and a temperature sensor under stimuli are shown in panels (ii) and (iii), respectively. (**I**) Pulse train output (bottom) from a pressure sensor–RO-ED (five-stage) system during a pressing-releasing cycle (top). (**J**) Output frequency of the sensor-RO-ED system under different pressures from data shown in (I). (**K**) Comparison of previously reported intrinsically stretchable circuits with our RO-ED circuit in terms of the integration scale level (number of transistors and logic gates). The different shapes and patterns of the symbols are used to mark the number of transistors involved in circuits from previous works, which are indicated by the reference numbers. The dashed line represents the trend of the degree of integration of stretchable electronics over time. All circuits were operated under V_{DD} = 5 V and V_{SS} = −5 V. [Photos in (D) and (H) were taken by J.-C.L.].

With good solvent resistance and patternability, our high-κ, trilayer dielectric is an important development for the scalable and transfer-free fabrication of low-voltage-driven intrinsically stretchable transistor arrays (Fig. 2Aii) (see materials and methods and figs. S11 to S16). The resulting arrays showed stretchable transistor properties, including 3-V operation with an on-off current ratio of ~10^6, negligible hysteresis (~20 mV under 3-V operation), low gate leakage (~10 pA), high mobility high yield (generally ~90%, maximum 100%

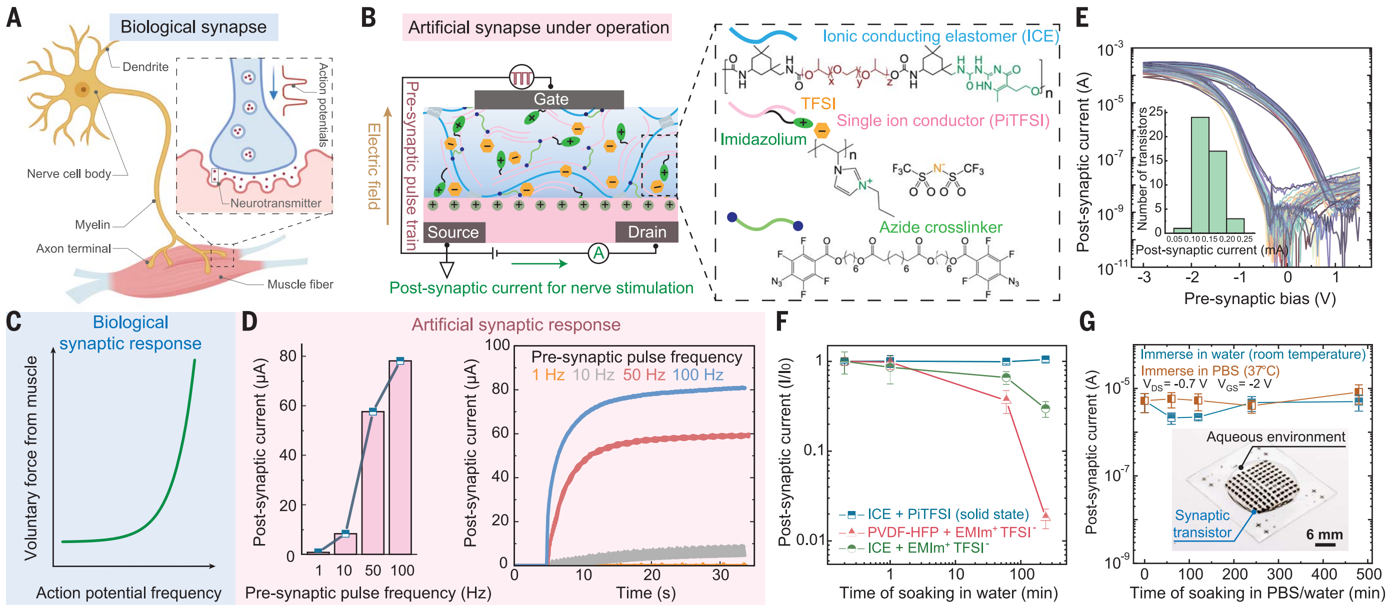


Fig. 4. All-solid-state soft artificial synapse for actuating downstream body movements based on gate input frequencies. (**A** and **B**) Comparison of the working mechanism of a biological synapse (A) with that of an artificial synaptic transistor (B). Components of the biological synapse are labeled. Material designs and selections of our artificial synapse are shown in the dashed box in (B). A, current meter. [Illustration was created with BioRender.com.] (**C**) In a biological synapse, larger muscle forces (output amplitude) are generated when action potentials with higher frequencies are transmitted through the motor nerve (*38*). (**D**) Correlation between presynaptic pulse-train frequency and postsynaptic current amplitude from a synaptic transistor. When gated by pulse-train signals (which mimic action potentials), mobile ions (which mimic neurotransmitters) migrate to the dielectric-semiconductor interface and induce a conducting channel. Higher frequency (together with a higher duty cycle) leads to higher postsynaptic current amplitudes. (**E**) Transfer curves and postsynaptic current distribution (inset) of a synaptic transistor array. Channel length and width are 80 and 320 μm, respectively; the dielectric thickness is 2 μm. (**F**) Comparison of the water stability of different dielectrics. Ionic conducting elastomer (ICE) plus polyelectrolyte (PiTFSI) shows stable device performance after 4 hours of immersion in water. Three transistors were monitored, and their postsynaptic current values were averaged. EMIm TFSI, 1-ethyl-3-methylimidazolium TFSI; PVDF-HFP, poly(vinylidene fluoride-co-hexafluoropropylene). (**G**) Stability test results using data averaged from three monitored synaptic transistors. The transistors were soaked in water or 37°C phosphate-buffered saline (PBS) solution for up to 8 hours. The inset is a photo showing a stretchable synaptic transistor array immersed in an aqueous environment. All error bars represent standard deviations. Postsynaptic currents were obtained from either the peak current of the pulse gated output for (D) after 30 s of operation or the transfer curve output at the highest presynaptic bias voltage for (F) to (G). [Photo in (G) was taken by J.C.-L.]

for two batches of transistor arrays containing 50 devices each; see materials and methods), high stretchability (up to 100% strain), and high cyclability (more than 1000 stretch-release cycles under 60% strain) (Fig. 2A iii and figs. S17 to S22). Compared with previously reported low-κ, stretchable dielectrics, our transconductance normalized by channel width (G_m/W_{ch}) was improved by >10-fold (Fig. 2G). Moreover, we achieved a subthreshold swing (~85 mV per decade) for stretchable transistors comparable to that of rigid polycrystalline Si (poly-Si) transistors (Fig. 2G and figs. S23 and S24). As a result, our stretchable transistor arrays showed a high on-off current ratio at a low gate voltage (greater than 10^4 under 0.75 V) (fig. S25 and table S2). The improvements in the key figures of merit for transistor performance (driving voltage, subthreshold swing, transconductance, on-off ratio, and leakage current) allow our soft transistors to operate with low power, which is ~100-fold less for both dynamic (1.7 pJ) and static (0.25 pW) power dissipations than that for existing stretchable transistors (Fig. 2H and fig. S26). Finally, we demonstrated the operation of a stretchable functional circuit (e.g., Pseudo-E inverter circuit) with a driving voltage of ±0.5 V under 100% strain (figs. S27 and S28) and a stretchable single-stage Pseudo-D inverter circuit with a gain value as high as 120 (fig. S29).

Signal conditioning circuit system for nerve-like pulse-train generation

By combining the optimal characteristics of individual transistors, the transistors can be further integrated to make low-voltage functional circuits for direct on-skin operation that mimic the sensory functions of biological cutaneous receptors (figs. S30 and S31) (*1*, *3*). During natural perception processes, somatosensory receptors transform stimulus input amplitudes (reception phase) into frequency-modulated pulse trains with a constant amplitude (encoding phase) for efficient and high-fidelity signal transmission (Fig. 3A). To recapitulate this process, we developed a circuit system with sensors that collect external stimuli, ring oscillators (ROs) for frequency modulation of sensor signals, and edge detectors (EDs) for action potential–like pulse-train signal generation (Fig. 3, B to D).

Typically, for sensor-RO, frequency-modulation systems, common designs involve inserting the sensor either between all stages of inverters and their positive power voltage (V_{DD}) (*10*, *11*) to tune the effective V_{DD} or between any two stages of inverters in the loop to introduce an additional resistor-capacitor (RC) delay (figs. S32 and S33). However, both methods would introduce large variations in the oscillation amplitudes with a limited dynamic range for frequency tunning. Conversely, frequency modulations in biological systems are amplitude decoupled and generally range from 0 to 100 Hz (*36*). To address this issue, we introduce a design that places the sensor into the second stage of the inverter (termed “sensing inverter”) (Fig. 3E). Based on the circuit modeling, the pull-up charging current of the sensing inverter can be effectively adjusted by the sensor response, whereas the operation of other stages is almost unaltered (fig. S34). Additionally, as a benefit of direct charging from the V_{DD}, the output of the sensing inverter can

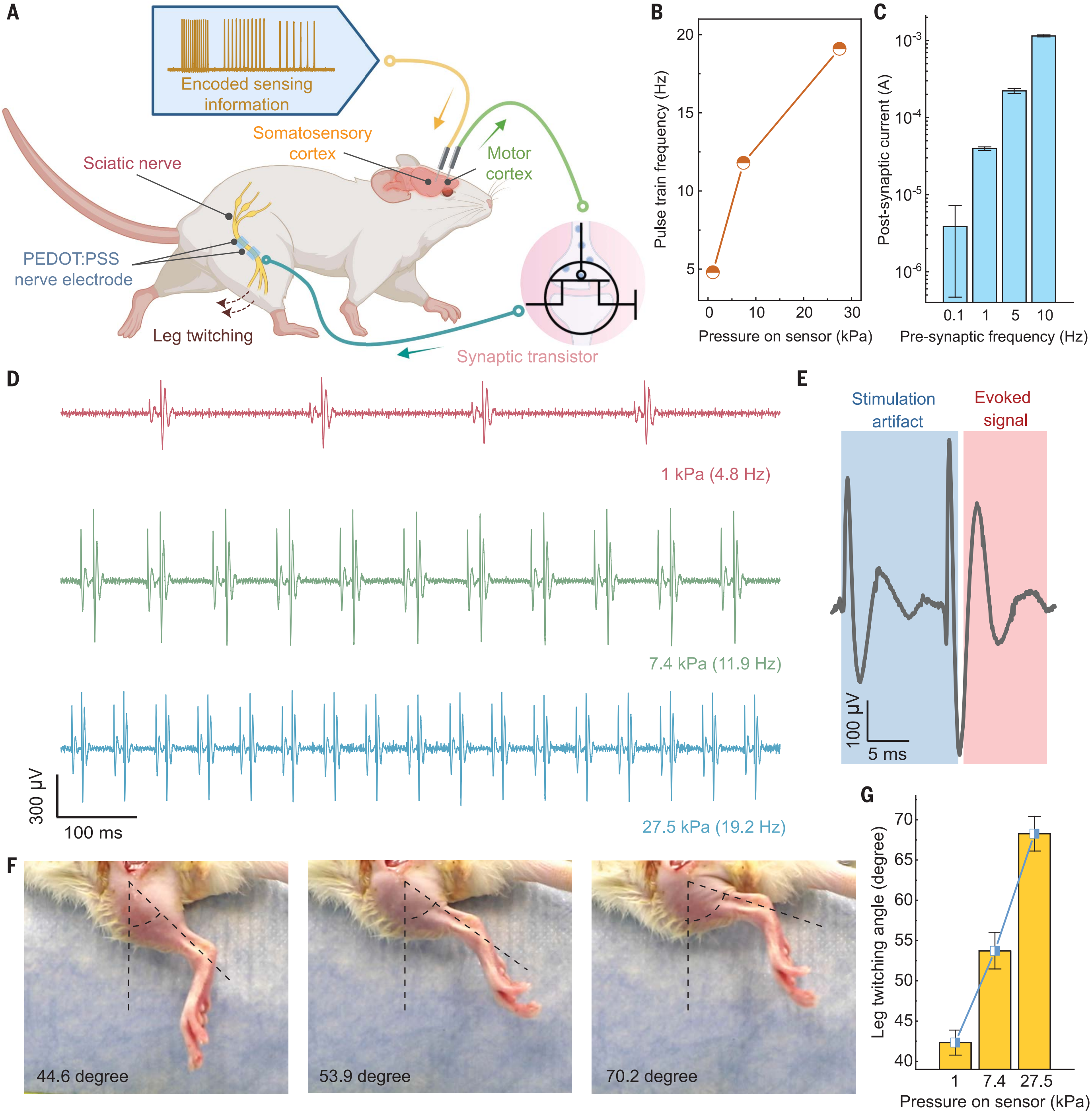


Fig. 5. Low-voltage-driven soft e-skin system for an artificial perception-actuation loop. (**A**) Schematic diagram showing the structure of the artificial sensorimotor system. A pressure sensor–RO-ED system was used to translate the applied force into pulse-train signals to stimulate the somatosensory cortex and evoke responses at the motor cortex. The processed and amplified signals recorded from the motor cortex were used as the gate input of the synaptic transistor, leading to different postsynaptic currents for hindlimb controls. [Illustration was created with BioRender.com.] (**B**) Pulse-train output frequencies under different pressure levels applied on a pressure sensor. Data points were obtained from the frequencies of the signal recorded at the motor cortex. (**C**) Synaptic transistor current output biased by different gating signal frequencies. Postsynaptic current was obtained by averaging three consecutive peak currents during biasing. (**D** and **E**) Recorded signals from the motor cortex corresponding to different pressures applied to the sensor. The results were filtered by a high-pass Butterworth filter with a cut-off frequency of 120 Hz. (**F**) Photos showing leg twitching in a representative animal model in response to different frequency stimulations correlating to different pressure inputs. (**G**) Correlation between leg twitching angle and applied pressure, which is calculated based on the stimulation frequency used. Data points represent average values from four rats. All error bars refer to standard deviations. [Photos in (F) were taken by W.W. and Y.J.]

maintain a large swing over 2.9 V even with a loading resistor of 1 gigohm (only 0.2 V with a loading resistor of 100 megohms for the existing design when that sensor was loaded within the RO's loop) (figs. S33 and S34). Thus, the large output signal from the sensing inverter can be gradually amplified by the inverters after the sensing inverter and stabilized around V_{DD} at the final RO output terminal, regardless of the sensor response. Furthermore, the combination of the large changes in the pull-up current and the unattenuated voltage amplitude enable a wide dynamic range of frequency change based on the sensor value (fig. S35). Experimentally, with our new design and the high on-off current ratios of our transistors, we realized a modulation frequency range that is relevant for physiological signals (i.e., 0 to 100 Hz) (*36*), along with a stable oscillation amplitude when the loading resistance varies (i.e., sensor response) between 0 and 5 gigohm (Fig. 3F and fig. S36). We also observed that the amplitude-decoupled frequency-modulation property could be well maintained even under 100% strain (figs. S37 and S38). Finally, to implement the multimodal sensory capability, we fabricated both five- and seven-stage ROs with different oscillation frequency ranges that can be used to distinguish information from multiple sensors (figs. S39 and S40).

The output waveforms of ROs alone would have a wide pulse duration with stage-dependent variations, which are different from natural action potentials with transient spikes and fixed narrow pulse widths (*1*, *36*). To better mimic the biological encoding process of pulse trains, we further developed ED circuits to "reshape" the RO output signals. Through a properly designed delay network and AND gate, the ED can effectively capture the rising edges of the incoming signals to generate sharp pulses (Fig. 3Gi). Here, two different delay network designs were investigated—the RC network and the inverter chain. We observed that the three-stage inverter chain–based ED showed a stable output amplitude under different input frequencies, whereas the RC network–based ED showed decreased output amplitudes with higher frequencies of the input signals, which is due to the roll-off behavior of the RC network (figs. S41 and S42). With the optimal geometry design of the delay network, the AND gate, and the corresponding circuit fabrication, we achieved a stretchable ED that can generate stable pulse signals with a duration of ~4 ms and an amplitude of ~5 V with different frequencies of square-wave input (up to 50 Hz) (Fig. 3G and figs. S43 and S44). This pulse width can be independently tuned through the stage number of the inverter chain or transistor size. Notably, with a proper geometry design, our ED can capture the rising edges of sinusoidal signals as well, making it possible to reshape the outputs from ROs with high oscillation frequencies that are unable to generate square wave–like signals (fig. S45). The function of our stretchable ED can be well maintained up to 100% strain with a minimal increase in the pulse width (figs. S46 and S47). We also note that pulses generated by piezoelectric, triboelectric, or ionic sensors are only due to repetitive pressing (*23*). These devices are not yet capable of encoding amplitude information to frequency. They also cannot properly tune the pulse duration and amplitude without signal conditioning circuits.

Using all the developed components, we proceeded to fabricate a monolithic integrated e-skin patch that is as soft as human skin (fig. S48) to emulate the processes of natural sensation with a driving voltage less than ±5 V (Fig. 3, B and C). By loading standard resistors ranging from 10 megohm to 5 gigohm (to mimic the sensing response range), the frequencies of the output pulse-train signals are modulated from 31 to 1 Hz for a seven-stage RO-based system and from 42 to 2 Hz for a five-stage system (figs. S49 and S50). Besides the steady inputs, our e-skin can perceive simulated vibration stimuli up to 20 Hz (fig. S51). As a benefit of the low subthreshold swing value and high mobility, this neuromorphic circuit can successfully output pulse-train signals with various frequencies, driven by ±0.75 V (fig. S52). Furthermore, carbon nanotube–based stretchable pressure sensors with three-dimensional pyramidal structures and thin-film temperature sensors were developed and integrated into the system to mimic natural mechanoreceptors and thermoreceptors, respectively (Fig. 3H and figs. S53 and S54). When applying pressure from 0 to 50 kPa or changing the temperature from 22° to 90°C, pulse-train signals were generated and fired faster in response to the pressure and temperature stimuli levels (Fig. 3, I and J, and figs. S55 and S56). The above e-skin circuit consisting of 54 stretchable transistors realized medium-scale integration for stretchable organic electronics (Fig. 3, C and K). Moreover, multimodal perception was implemented by dividing the pulse-train signals of different stimuli into distinct frequency regimes (figs. S57 and S58). When continuously biased under ±5 V under ambient conditions, the system was observed to remain functional after 8 hours (fig. S59). However, the development of stretchable encapsulation materials and stable organic semiconductors is still needed for invariant oscillation during continuous operation (figs. S60 and S61).

All-solid-state artificial synapses to trigger downstream actuations

A characteristic of the natural sensory feedback loop is bidirectional signal communication between perception and actuation (*9*, *37*). To complete the sensorimotor loop, in addition to the soft e-skin system for sensation, we next addressed the need to pass the frequency-encoded sensor information to the central nervous system and to actuate downstream muscle responses based on the input stimulus. The challenge is to appropriately translate the frequency-modulated signals to various levels of body movements (Fig. 4, A and C) (*38*). To this end, synaptic transistors based on ionic gating are particularly promising because they have a working mechanism similar to that of biological synapses (Fig. 4, B and D) (*39*). However, previously reported synaptic transistors were based on either liquid or gel components with poor environmental stabilities or rigid materials with limited tissue conformability (*10*, *40*). In this work, we developed a solid-state ionic dielectric by blending a previously designed ionic conducting elastomer (ICE) with high ion conductivity and a single-ion conductive polyelectrolyte {poly[(1-vinyl-3-propyl-imidazolium) bis(trifluoromethanesulfonyl)imide] (PiTFSI)} (Fig. 4B and figs. S62 and S63; see materials and methods). Compared with small-molecule ionic species that suffered from dissolution in biofluids, as detailed in previous reports, a polyelectrolyte can be stably fixed in the elastomeric matrix. Using azide cross-linkers, we patterned this ionic dielectric and fabricated an all-solid-state stretchable synaptic transistor array that has good uniformity (Fig. 4E and fig. S63). The synaptic behavior was observed to maintain the large hysteresis window from the transfer curves under 50% strain (fig. S64). Notably, when presynaptic pulses with different frequencies were used as gate inputs to the synaptic transistor, we observed large changes in the amplitudes of postsynaptic currents, that is, a ~4-order-of-magnitude output difference between 1 and 100 Hz inputs and a ~7-order-of-magnitude output difference when extending to 800 Hz (Fig. 4D and fig. S65). This is an important feature that is necessary in order to use frequency-encoded pressure amplitude information to generate different levels of body movements through the amplitudes of postsynaptic currents. Furthermore, after soaking in physiological fluids up to 8 hours, our all-solid-state artificial synapse showed a substantially reduced variation in postsynaptic current, whereas those containing small molecular ionic liquid lost ~98.2% of the output current (Fig. 4, F and G, and figs. S66 and S67).

Demonstration of the sensorimotor loop with monolithic soft e-skin

With all essential components in the e-skin system reproducing the biological perception-actuation loop, we proceeded to test the utility of this platform in a live rat model (Fig. 5, A to C). Briefly, we first connected our soft e-skin to the rat somatosensory cortex to emulate skin

sensation, which is expected to trigger feedback responses at the motor cortex. Evoked motor signals were then passed through the artificial synapse to stimulate the sciatic nerve for downstream muscle actuation, thus completing the artificial sensorimotor loop. Experimentally, when the sensor was subject to forces of different magnitudes, we observed that the digitalized inputs for the somatosensory cortex could successfully evoke a response at the motor cortex (Fig. 5, D and E, and fig. S68). After further amplification and signal processing, with the recorded evoked signals serving as the gate input of the synaptic transistor, the amplitudes of the postsynaptic currents were found to scale with applied pressure. The synaptic channel was interfaced with the sciatic nerve through a low-impedance poly(3,4-ethylenedioxythiophene):polystyrene sulfonate (PEDOT:PSS) electrode (*26*), resulting in the contraction of the biceps femoris muscle (see materials and methods). Similar to the natural sensory feedback process that elicits more intense reactions to stronger force stimuli, our bidirectional e-skin system also elicits larger leg twitching angles with respect to the increased pressure inputs (Fig. 5, F and G, and figs. S69 and S70). This demonstrates the potential feasibility of artificial e-skin for neuroprosthetics. This e-skin patch was attached on rat skin for 30 hours, and no skin irritation was observed (fig. S71).

Through rational material design and device engineering, we achieved a monolithically integrated, soft e-skin system without rigid electronic components that features a low driving voltage, a high circuit complexity, and biomimetic sensory feedback functions. Our described neuromorphic system combines all the desired electrical and mechanical features of skin in one single device platform. This advance will help pave the way for next-generation prosthetic skins, human-machine interfaces, and neurorobotics.

REFERENCES AND NOTES

1. A. Handler, D. D. Ginty, *Nat. Rev. Neurosci.* **22**, 521–537 (2021).
2. R. S. Johansson, J. R. Flanagan, *Nat. Rev. Neurosci.* **10**, 345–359 (2009).
3. L. E. Osborn *et al.*, *Sci. Robot.* **3**, eaat3818 (2018).
4. L. R. Hochberg *et al.*, *Nature* **485**, 372–375 (2012).
5. G. Valle *et al.*, *Neuron* **100**, 37–45.e7 (2018).
6. S. N. Flesher *et al.*, *Science* **372**, 831–836 (2021).
7. K. Lee *et al.*, *Nat. Biomed. Eng.* **4**, 148–158 (2020).
8. M. Wang *et al.*, *Adv. Mater.* **33**, e2003014 (2021).
9. A. Chortos, J. Liu, Z. Bao, *Nat. Mater.* **15**, 937–950 (2016).
10. Y. Kim *et al.*, *Science* **360**, 998–1003 (2018).
11. B. C. K. Tee *et al.*, *Science* **350**, 313–316 (2015).
12. Y. H. Dai, H. W. Hu, M. Wang, J. Xu, S. H. Wang, *Nat. Electron.* **4**, 17–29 (2021).
13. J. Lee *et al.*, *Nat. Electron.* **4**, 291–301 (2021).
14. I. You *et al.*, *Science* **370**, 961–965 (2020).
15. M. L. Hammock, A. Chortos, B. C. Tee, J. B. Tok, Z. Bao, *Adv. Mater.* **25**, 5997–6038 (2013).
16. F. Liu *et al.*, *Sci. Robot.* **7**, eabl7344 (2022).
17. T. Someya, Z. Bao, G. G. Malliaras, *Nature* **540**, 379–385 (2016).
18. J. A. Rogers, T. Someya, Y. Huang, *Science* **327**, 1603–1607 (2010).
19. H. Fang *et al.*, *Nat. Biomed. Eng.* **1**, 1–12 (2017).
20. D. Khodagholy *et al.*, *Nat. Commun.* **4**, 1575 (2013).
21. K. Takei *et al.*, *Nat. Mater.* **9**, 821–826 (2010).
22. T. Someya *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 9966–9970 (2004).
23. Y. Dobashi *et al.*, *Science* **376**, 502–507 (2022).
24. H. Tan *et al.*, *Nat. Commun.* **11**, 1369 (2020).
25. S. Lee *et al.*, *Science* **370**, 966–970 (2020).
26. Y. Jiang *et al.*, *Science* **375**, 1411–1417 (2022).
27. J. Xu *et al.*, *Nat. Mater.* **18**, 594–601 (2019).
28. X. Ji *et al.*, *Nat. Commun.* **12**, 2480 (2021).
29. Y. Q. Zheng *et al.*, *Science* **373**, 88–94 (2021).
30. S. Wang *et al.*, *Nature* **555**, 83–88 (2018).
31. B. Wang *et al.*, *Chem. Rev.* **118**, 5690–5754 (2018).
32. F. A. Viola *et al.*, *Nat. Commun.* **12**, 5842 (2021).
33. M. Halik *et al.*, *Nature* **431**, 963–966 (2004).
34. A. K. Jonscher, *J. Phys. D Appl. Phys.* **32**, R57–R70 (1999).
35. H. Sirringhaus, *Adv. Mater.* **17**, 2411–2425 (2005).
36. D. R. Lesniak *et al.*, *eLife* **3**, e01488 (2014).
37. K. I. Song *et al.*, *Nat. Commun.* **11**, 4195 (2020).
38. D. Purves *et al.*, *Neurosciences* (De Boeck Supérieur, ed. 6, 2019).
39. Y. Lee *et al.*, *Sci. Adv.* **4**, eaat7387 (2018).
40. F. Molina-Lopez *et al.*, *Nat. Commun.* **10**, 2676 (2019).
41. J. Xu *et al.*, *Science* **355**, 59–64 (2017).
42. J. Y. Oh *et al.*, *Nature* **539**, 411–415 (2016).
43. H. Ren *et al.*, *ACS Appl. Electron. Mater.* **3**, 1458–1467 (2021).
44. Y. Zheng *et al.*, *Adv. Funct. Mater.* **29**, 1905340 (2019).
45. J. Liu *et al.*, *Nat. Commun.* **11**, 3362 (2020).
46. Y. Zheng *et al.*, *Nat. Commun.* **12**, 5701 (2021).
47. A. Chortos *et al.*, *Adv. Mater.* **26**, 4253–4259 (2014).
48. C. Lu, W. Y. Lee, C. C. Shih, M. Y. Wen, W. C. Chen, *ACS Appl. Mater. Interfaces* **9**, 25522–25532 (2017).
49. J. Y. Oh *et al.*, *Sci. Adv.* **5**, eaav3097 (2019).
50. C. Zhu, H. C. Wu, G. Nyikayaramba, Z. A. Bao, B. Murmann, *IEEE Electron Device Lett.* **40**, 1630–1633 (2019).
51. G. Y. Zhang *et al.*, *Chem. Mater.* **29**, 7645–7652 (2017).
52. Y. L. Rao *et al.*, *J. Am. Chem. Soc.* **138**, 6020–6027 (2016).
53. C. C. Shih, W. Y. Lee, C. Lu, H. C. Wu, W. C. Chen, *Adv. Electron. Mater.* **3**, 1600477 (2017).
54. B. Kang *et al.*, *Chem. Mater.* **30**, 6353–6360 (2018).
55. W. C. Wang *et al.*, *Nat. Electron.* **4**, 143–150 (2021).
56. H. Ren *et al.*, *J. Mater. Chem. C Mater. Opt. Electron. Devices* **8**, 15646–15654 (2020).
57. L. Cai, S. Zhang, J. Miao, Z. Yu, C. Wang, *ACS Nano* **10**, 11459–11468 (2016).
58. C. X. Zhu *et al.*, *Nat. Electron.* **1**, 183–190 (2018).
59. K. Sim *et al.*, *Sci. Adv.* **5**, eaav5749 (2019).

ACKNOWLEDGMENTS

We thank A. Nepveux for helping with experiment preparation and Asahi Kasei Co. for providing SEBS polymer. **Funding:** This work was supported in part by the Stanford Wearable Electronics Initiative (eWEAR), Stanford SystemX Alliance, and the Wu Tsai Neuroscience Institute. Part of this work was performed at the Stanford Nano Shared Facilities (SNSF), which is supported by the National Science Foundation under award ECCS-2026822. Y.-H.K. acknowledges support from National Research Foundation grant 2021R1A2B5B03086367. **Author contributions:** W.W., Y.J., D.Z., Z.Z., and Z.B. designed the project and experiments. W.W., Y.J., D.Z., and Z.Z. carried out experiments and collected related data. W.W. carried out trilayer dielectric and low-voltage-driven transistor design and fabrication. D.Z. and W.W. carried out circuit designs. W.W. carried out circuit fabrication and characterizations. S.N. assisted with circuit analyses and measurements. D.Z. carried out transistor modeling and circuit simulations. Y.J., D.L., and W.W. designed and carried out the e-skin system demonstration on rats. W.W., Y.J., and Z.Z. designed the all-solid-state synaptic transistor. S.C. designed and synthesized the polyelectrolyte. X.Y. designed and synthesized the ionic conducting elastomer. Y.Z. designed and synthesized the azide compound. Y.-H.K. and J.K. designed and synthesized the conjugated polymer (29-DPP-SVS). W.W. and Z.Z. carried out thickness measurements. J.-C.L. conducted tensile testing. H.G. performed x-ray photoelectron spectroscopy characterization. Y.J. and Q.L. carried out atomic force microscopy measurements. H.L. carried out focused ion beam scanning electron microscopy measurements. J.-C.L., R.N., C.X., X.J., and D.Z. assisted with device photography. Z.Z., Y.-X.W., C.Z., C.-C.S., and Y.N. helped with experimental design and preparation. W.W., Y.J., D.Z., Z.Z., J.B.-H.T., and Z.B. wrote the manuscript. All authors reviewed and commented on the manuscript. **Competing interests:** Stanford University has filed a patent application related to this technology. The patent application number is 63/457550. **Data and materials availability:** All data are available in the main text or the supplementary materials.

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ade0086
Materials and Methods
Figs. S1 to S71
Tables S1 and S2
References (*60–93*)
MDAR Reproducibility Checklist

Submitted 19 July 2022; accepted 31 March 2023
10.1126/science.ade0086

GLOBAL HYDROLOGY

Satellites reveal widespread decline in global lake water storage

Fangfang Yao[1]*, Ben Livneh[1,2], Balaji Rajagopalan[1,2], Jida Wang[3], Jean-François Crétaux[4], Yoshihide Wada[5,6], Muriel Berge-Nguyen[4]

Climate change and human activities increasingly threaten lakes that store 87% of Earth's liquid surface fresh water. Yet, recent trends and drivers of lake volume change remain largely unknown globally. Here, we analyze the 1972 largest global lakes using three decades of satellite observations, climate data, and hydrologic models, finding statistically significant storage declines for 53% of these water bodies over the period 1992–2020. The net volume loss in natural lakes is largely attributable to climate warming, increasing evaporative demand, and human water consumption, whereas sedimentation dominates storage losses in reservoirs. We estimate that roughly one-quarter of the world's population resides in a basin of a drying lake, underscoring the necessity of incorporating climate change and sedimentation impacts into sustainable water resources management.

Lakes cover 3% of the global land area (*1*), storing standing or slowly flowing water that provides essential ecosystem services of fresh water and food supply, waterbird habitat, cycling of pollutants and nutrients, and recreational services (*2*). Lakes are also key components of biogeochemical processes and regulate climate through cycling of carbon (*3*). The potential goods and services from lakes are modulated by lake water storage (LWS) (*4*), which fluctuates in response to changes in precipitation and river discharge, as well as in response to direct human activities (damming and water consumption) and climate change (*5*–*7*). It is well documented that some of the world's largest lakes have recently experienced a decline in water storage (*4*, *6*–*9*). However, the drivers of LWS decline have either been poorly constrained by inconsistent methods and assumptions or remain unknown for the majority of unstudied large lakes (>100 km^2) that lack a decadal-scale LWS record (*10*). For example, recent level declines in the Caspian Sea have been primarily attributed to entirely different processes, either evaporative losses from the water body (*11*) or decreasing river discharge (*12*). Similarly, the recent decadal decline of China's largest freshwater body, Lake Poyang, has been separately attributed to either the operation of the Three Gorges Dam (*9*) or climate variability (*13*). More broadly, there have been indications of global shifts in LWS over recent decades. Between 1984 and 2015, a loss of 90,000 km^2 of permanent water area was observed by satellites—an area equivalent to the surface of Lake Superior, whereas 184,000 km^2 of new water bodies, primarily reservoirs, were formed elsewhere (*14*). Yet, trends and drivers of global LWS remain poorly known, which impedes sustainable management of surface water resources, both now and in the future.

The estimation of trends and variability in global LWS has been complicated by modeling and observational limitations. Current global hydrologic models either neglect LWS changes (*15*) or provide oversimplified simulations using one-dimensional models of lake volume changes (*16*). In situ measurements of lakes are spatially sparse, have irregular temporal coverage, or are generally in decline (*17*). Satellites provide a crucial dimension for assessing large-scale LWS variability through repeated observations of lake areas and water levels from space. However, sensor limitations, such as coarse resolution, infrequent overpasses, large intertrack spacing, and mission gaps, prevent the direct development of a global inventory of LWS changes over time. As a result, existing global-scale studies that document lake volume changes lack the capability to attribute decadal-scale LWS variability because of limited spatial coverage (*10*, *18*, *19*), short temporal duration (<2 years) (*20*), or large gaps in the LWS time series (>9 years) (*21*). Using NASA's ICESat-2 satellite and a one-dimensional model (assuming a constant lake area), a recent study mapped water levels and storages in 227,386 global water bodies over 2018–2020 and found that reservoirs, defined as water bodies regulated by a dam, dominated seasonal variability in global lake water storage (*20*). Given the brevity of the study period (<2 years) and limited attribution, i.e., only separating natural lakes and reservoirs (*20*), decadal-scale LWS variability and attribution remain an open question. Another study combined water levels from ICESat and ICESat-2 to map LWS changes in 6567 lakes over 2003–2020 and compared LWS changes across different climate regimes (*21*) but suffered from a 9-year discontinuity (2010–2018) in LWS time series and limited ability to capture and diagnose drivers of interannual variability and trends over the recent decades. Therefore, the human and climate change footprints on global LWS changes over decadal timescales remain critically unknown.

To address this challenge, we construct a global database of time-varying LWS (GLWS) from 1992 to 2020, and then decouple the impacts of anthropogenic and natural factors on decadal-scale variability in LWS (Fig. 1). This GLWS archive consists of subyearly storage time series for 1972 large water bodies, including 1051 natural lakes (100 to 377,002 km^2) and 921 reservoirs (4 to 67,166 km^2), which account for 96 and 83% of Earth's natural lake and reservoir storage, respectively (*1*, *22*). We focus on large lakes because of the fidelity of satellite observations at this scale and their dominance in controlling total lake volume change (*8*, *20*), as well as their importance for human and wildlife populations (*23*). We leverage recent advances in both algorithms (*24*) and cloud-based parallel computing to construct time-varying water areas for these lakes using a total of 248,649 Landsat satellite images at a 30-m resolution (materials and methods). We estimate lake volume variability by combining water areas with water surface elevation measurements from satellite altimeters, including CryoSat-2, ENVISAT, ICESat, ICESat-2, Jason series 1-3, SARAL, and Sentinel 3. On average, we derive six estimates per year over the 28-year study period for each studied water body. We further provide a global-scale attribution of volume trends in natural lakes using a statistical-learning framework that incorporates major natural and anthropogenic drivers estimated from global climate data and hydrologic models. For reservoirs, we aggregate impacts of recent dam construction and subsequent reservoir infilling using newly compiled global dam and reservoir inventories (*25*), as well as sedimentation using in situ sediment surveys and upscaling methods. Finally, we isolate lake storage trends between arid and humid regions and quantify the numbers of local population subject to lake water losses.

Global LWS trends and drivers

We identify widespread decline in global LWS over the past 28 years. Over half (53 ± 2%) of the large lakes experienced significant water losses

[1]Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado Boulder, Boulder, CO 80309, USA. [2]Department of Civil, Environmental and Architectural Engineering, University of Colorado Boulder, Boulder, CO 80309, USA. [3]Department of Geography and Geospatial Sciences, Kansas State University, Manhattan, KS 66506, USA. [4]Laboratoire d'Études en Géophysique et Océanographie Spatiales (LEGOS), Université de Toulouse, CNES-IRD-CNRS-UT3, Centre National d'Études Spatiales (CNES), 31013 Toulouse, France. [5]Climate and Livability Initiative, Center for Desert Agriculture, Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia. [6]International Institute for Applied Systems Analysis (IIASA), A-2361 Laxenburg, Austria.
***Corresponding author. Email: fangfang.yao@colorado.edu; yaoff.luke@gmail.com**

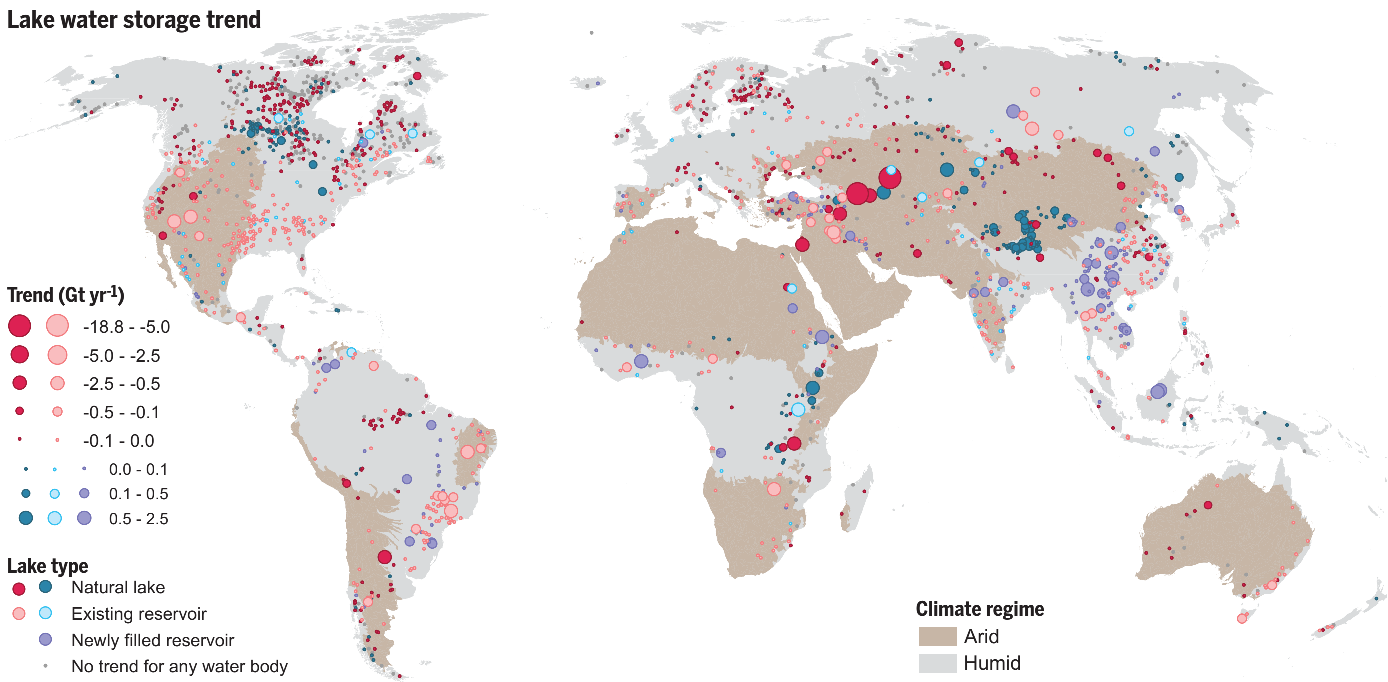


Fig. 1. Widespread storage decline in large global lakes from October 1992 to September 2020. Lake water storage (LWS) trends for 1058 natural lakes (dark red and dark blue dots) and 922 reservoirs (light red and light blue dots). Recently filled reservoirs after 1992 are denoted as light purple dots. All colored dots denote statistically significant trends ($p < 0.1$), whereas no significant trends are shown as gray dots. Classification of climate regimes between arid and humid regions was done using the aridity index [ratio of mean annual precipitation to mean annual potential evapotranspiration (materials and methods)].

($p < 0.1$) (Fig. 1). LWS loss prevails across major global regions including western Central Asia, the Middle East, western India, eastern China, northern and eastern Europe, Oceania, the conterminous United States, northern Canada, southern Africa, and most of South America. Roughly one-quarter (24%) of the large lakes experienced significant water gains, which are largely found in dam-construction hotbeds and in remote or underpopulated regions, such as the Inner Tibetan Plateau and the Northern Great Plains of North America. Globally, LWS showed a net decline at a rate of −21.51 ± 2.54 Gt year^{-1} (Fig. 2, A and D), or by 602.28 km^3 in accumulative volume—equivalent to the total water use in the US for the entire year of 2015 or 17 times the volume of Lake Mead, the largest reservoir in the United States (Fig. 2, A and D). The accumulative volume loss in global LWS is about 40% larger than the mean of annual variations (i.e., differences between maximum and minimum values) in global LWS over the period 1992–2020, masking its importance to the water cycle and freshwater resources.

Globally, natural lake volume declined at a net rate of −26.38 ± 1.59 Gt year^{-1}, of which 56 ± 9% is attributable to direct human activities and changes in temperature and potential evapotranspiration (PET), i.e., evaporative demand (Fig. 2G). A total of 457 natural lakes (43%) had significant water losses with a total rate of −38.08 ± 1.12 Gt year^{-1}, whereas significant water gains were found in 234 natural lakes (22%) at a total rate of 13.02 ± 0.41 Gt year^{-1}. The remaining 360 lakes (35%) showed no significant trends (Fig. 1). Over 80% of the total decline in drying lakes stems from the 26 largest losses (>0.1 Gt year^{-1}, $p < 0.1$) (Fig. 3A). In particular, the largest inland water body, the Caspian Sea, accounts for 49% of the total decline and 71% of the net decline in natural lake volume. Yet, excluding the Caspian Sea does not alter the dominance of anthropogenic impacts on the observed decline in natural LWS (fig. S1). Unsustainable water consumption dominates the observed drying of the Aral Sea (−6.59 ± 0.36 Gt year^{-1}) in Central Asia, Lake Mar Chiquita (−0.75 ± 0.09 Gt year^{-1}) in Argentina, the Dead Sea (−0.63 ± 0.04 Gt year^{-1}) in the Middle East, and the Salton Sea (−0.11 ± 0.01 Gt year^{-1}) in California. Increasing temperature and PET led to the complete disappearance of Lake Good-e-Zareh (−0.48 ± 0.17 Gt year^{-1}) in Afghanistan, Toshka lakes (−0.13 ± 0.10 Gt year^{-1}) in Egypt, and marked drying of Lake Kara-Bogaz-Gol (−1.27 ± 0.09 Gt year^{-1}) in Turkmenistan, Lake Khyargas (−0.35 ± 0.03 Gt year^{-1}) in Mongolia, and Lake Zonag (−0.26 ± 0.14 Gt year^{-1}) in China. The remaining change is primarily attributable to changes in precipitation and runoff, including the Caspian Sea (−18.80 ± 0.93 Gt year^{-1}), Lake Urmia (−1.05 ± 0.06 Gt year^{-1}) in Iran, the Great Salt Lake (−0.29 ± 0.08 Gt year^{-1}) in the United States, Lake Poyang (−0.13 ± 0.12 Gt year^{-1}) in China, Lake Titicaca (−0.12 ± 0.08 Gt year^{-1}) on the border of Bolivia and Peru, and others, which largely agrees with existing studies (*8*, *12*, *13*, *26*). Whereas a recent study suggests that the drying of Lake Urmia was primarily attributed to human activities (*27*), we find that naturalized flows explain 67% of the variance in the annual mean lake volume compared with 52% explained by human water consumption. Thus, the decline of Lake Urmia is likely a concurrent result of both reduced naturalized flows and human activities. Arctic lakes were mostly in decline owing to a combination of changes in precipitation, runoff, temperature, and PET (Fig. 3A), which are likely a concurrent result of natural variability and climate change. Globally, temperature-and-PET changes dominate water loss in 21% of the drying lakes (Fig. 3A). Approximately one-third of the total decline in all drying lakes is offset by lake storage increases elsewhere, largely in remote or sparsely populated areas such as the Inner Tibetan Plateau, Northern Great Plains, and Great Rift Valley (Fig. 3B). These storage increases were driven primarily by changes in precipitation and runoff, and to a lesser extent, by temperature-and-PET changes and reduced human water use. For example, Lake Sevan in Armenia experienced an increasing storage trend (0.20 ± 0.02) owing to the enforcement of conservation laws on water withdrawal since the 2000s (*28*).

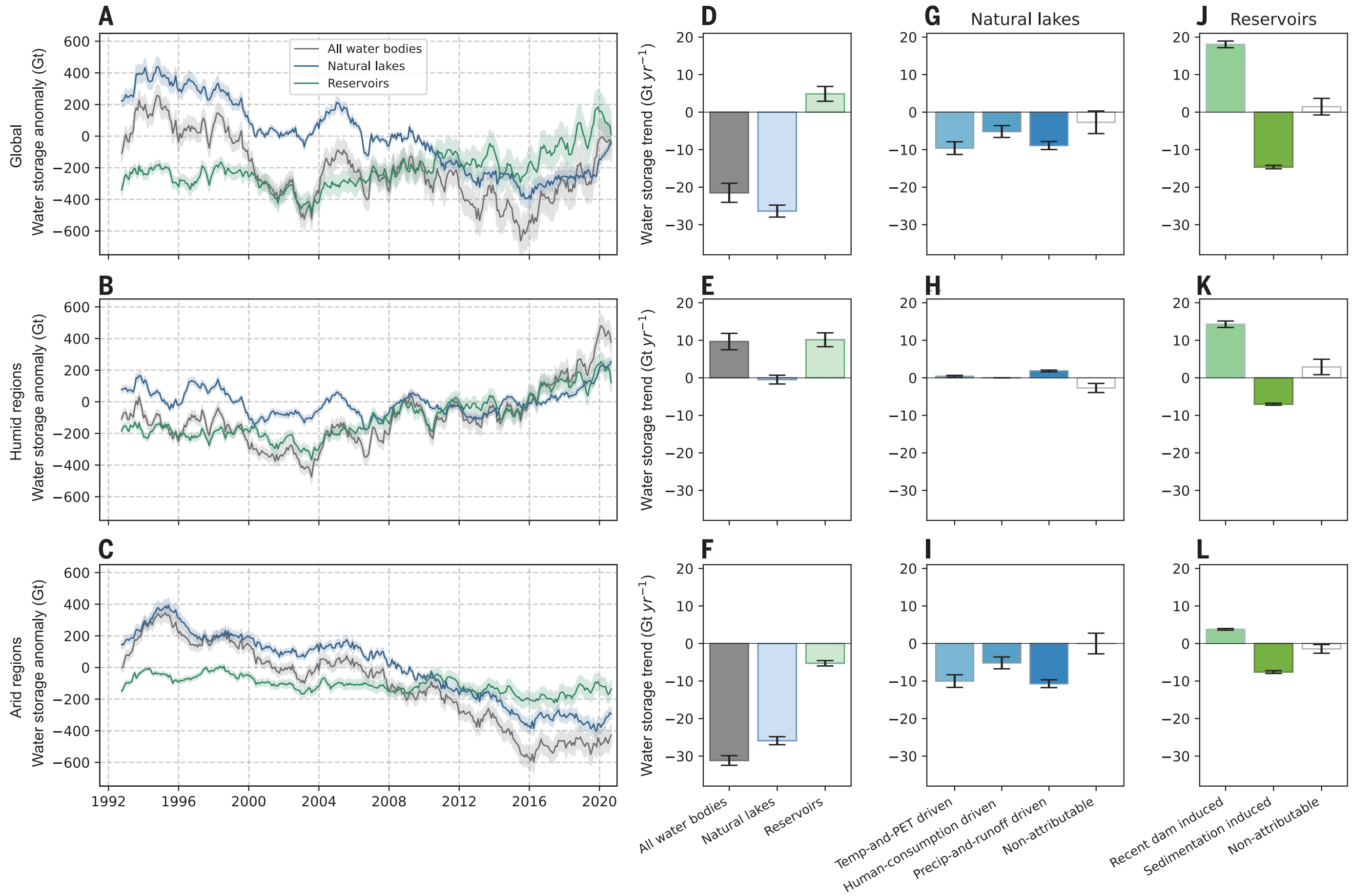


Fig. 2. Time series and drivers of global lake water storage change, October 1992 to September 2020. (**A** to **F**) Time series and trends of aggregate storage anomalies for each type of water body for global, humid, and arid regions, respectively. (**G** to **L**) Attribution of storage trends in natural lakes and reservoirs. Temp, Precip, and PET stand for temperature, precipitation, and potential evapotranspiration, respectively. The shading denotes the LWS uncertainties in all water bodies (gray shading), natural lakes (blue shading), and reservoirs (green shading) at a 95% confidence interval. The error bars show the aggregate uncertainty in LWS trends at a 95% confidence interval. For natural lakes, 57% of the net global decline is attributable to human activities and increasing temperature and potential evapotranspiration. Recent dam construction, largely in humid basins, supported the net increase in global reservoir storage, although more than 80% of the increased storage in recently filled reservoirs is offset by sedimentation-induced storage loss in existing reservoirs.

Nearly two-thirds (64 ± 4%) of all large reservoirs experienced significant storage declines, although reservoirs showed a net global increase at a rate of 4.87 ± 1.98 Gt year^{-1}, owing to 183 (20%) recently filled reservoirs. Storage declines in existing reservoirs, i.e., filled before 1992, were observed in most regions. Global storage decline in existing reservoirs (–13.19 ± 1.77 Gt year^{-1}) can be largely attributed to sedimentation (Fig. 2J), which is consistent with observations of substantial storage decline in US reservoirs due to sedimentation (*29*, *30*). Globally, sedimentation-induced storage loss offsets more than 80% of the increased storage from new dam construction (Fig. 2J). Our finding suggests that sedimentation is the primary contributor to the global storage decline in existing reservoirs and has a larger impact than hydroclimate variability, i.e., droughts and recovery from droughts (*19*). Recent droughts may have contributed to reservoir storage declines, particularly in the southwestern United States (*31*), eastern and southern Brazil (*32*), the Middle East (*33*), southern India (*34*), eastern and southeastern Asia (*35*), eastern Oceania (*36*), and most of Europe (*37*). However, drought impacts on reservoir LWS have been partially offset by wetting trends elsewhere (*12*), such as the headwaters of the Nile River, southeastern Canada, and Mexico (Fig. 1).

LWS trends in arid and humid regions

Arid regions experienced a net storage decline for both natural lakes and reservoirs at a rate of –25.90 ± 1.07 and –5.27 ± 0.71 Gt year^{-1}, respectively. Approximately 60% of the water bodies in arid regions had significant water losses. More than 70% of the total LWS loss in these water bodies stemmed from 10 basins with either a drying climate or unsustainable human water consumption or both, such as in the Aral Sea Basin and the Colorado River Basin (Fig. 4). The basins of the Caspian Sea, the Aral Sea, Lake Good-e-Zareh, Lake Urmia, Lake Rukwa, Lake Chiquita, the Dead Sea, and western Mongolia are large contributors to the global net decline in natural LWS. However, the vast majority (93%) of natural lakes with significant water losses are found in basins elsewhere and are distributed widely across the globe, which is in line with the widespread global decline in LWS seen in Fig. 1. In arid regions, roughly one-quarter of the significant water losses in natural lakes were dominated by changes in temperature and PET or human activities, and another 37% of the water losses were primarily attributable to reduced natural flows. Storage declined in more than two-thirds (68 ± 3%) of reservoirs in arid regions. The net reservoir storage loss was mostly attributable to sedimentation (Fig. 2L), although droughts likely aggravated reservoir storage losses, such

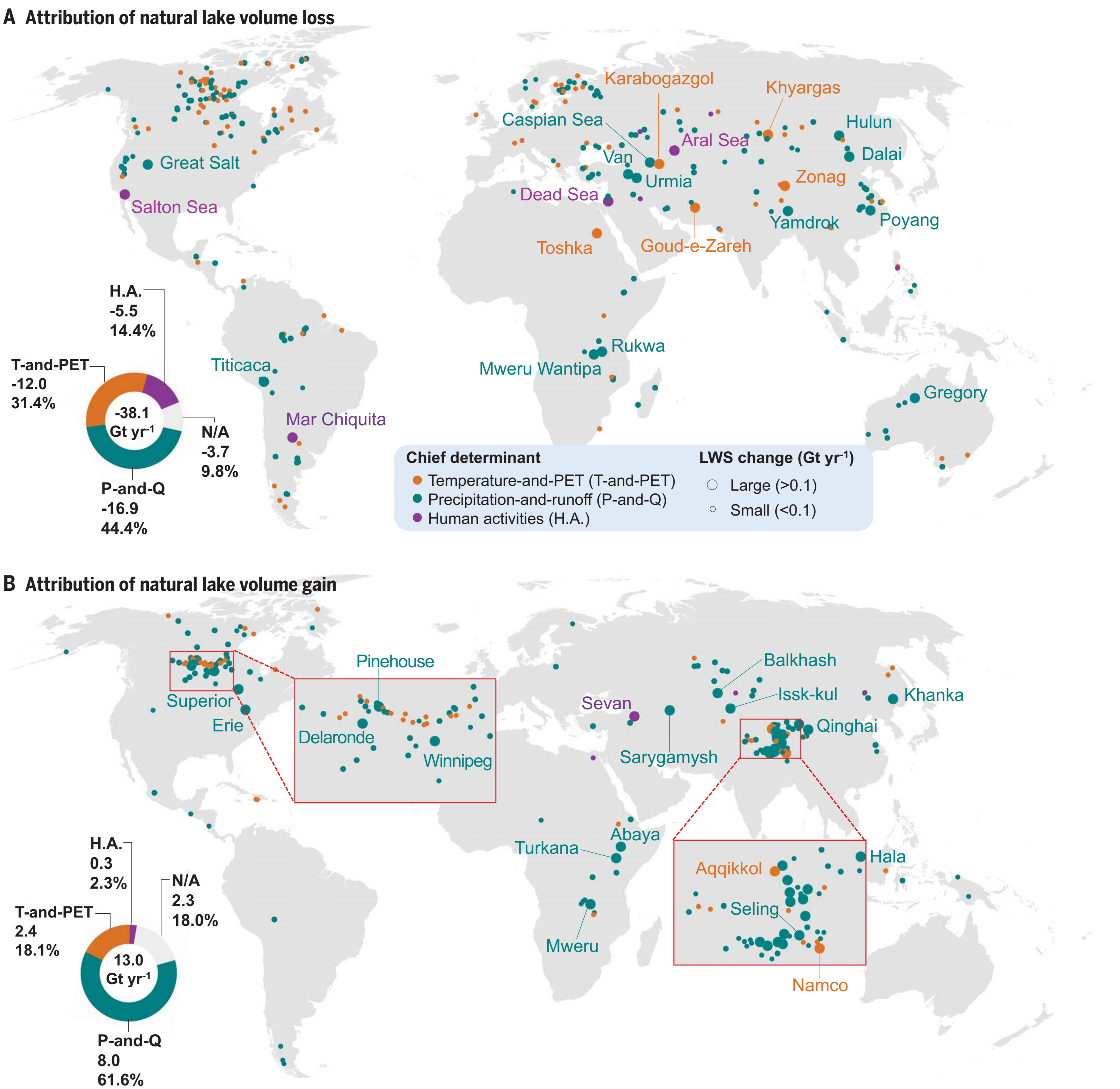


Fig. 3. Attributions of significant volume changes in natural lakes. (**A**) Chief determinants for volume losses. (**B**) Chief determinants for volume gains. PET, potential evapotranspiration. The inset pie charts show the aggregate impact (by magnitude) of each determinant on the basis of relative contributions. For clarity, lake volume changes that are not significantly attributable (N/A) are not shown in panels (A) and (B), but their proportions are included in the insets.

as in the Colorado River Basin and the Tigris and Euphrates Basin (*33*, *38*). Newly filled reservoirs resulted in a total of 3.77 ± 0.21 Gt year^{-1} of water gain, which partially offsets the net decline in reservoir storage by 42%. These new reservoirs are mainly found in the Nile, Volta, Yellow, and Amur River Basins. In the Nile Basin, newly filled reservoirs account for 41% of the increased basin water storage, and the remaining 61% was largely attributable to the rapid rise of the regulated Lake Victoria in the headwater of the White Nile under precipitation extremes in 2019–2020 (*39*). A few other arid basins, such as the Inner Tibetan Plateau, Lake Balkhash, Great Rift Valley, and basins, experienced major water gains due to recovery from droughts and increasing precipitation (*8*, *12*, *40*) (Fig. 3).

New construction of reservoirs drove a net LWS gain in humid regions at a rate of 9.66 ± 2.19 Gt year^{-1}, although 54% of the remaining water bodies showed significant storage declines. A total of 124 newly filled reservoirs in humid regions gained storage at a rate of 14.28 ± 0.86 Gt year^{-1}, which offsets the total storage decline in existing reservoirs (–4.15 ± 1.63 Gt year^{-1}). Globally, more than 85% of

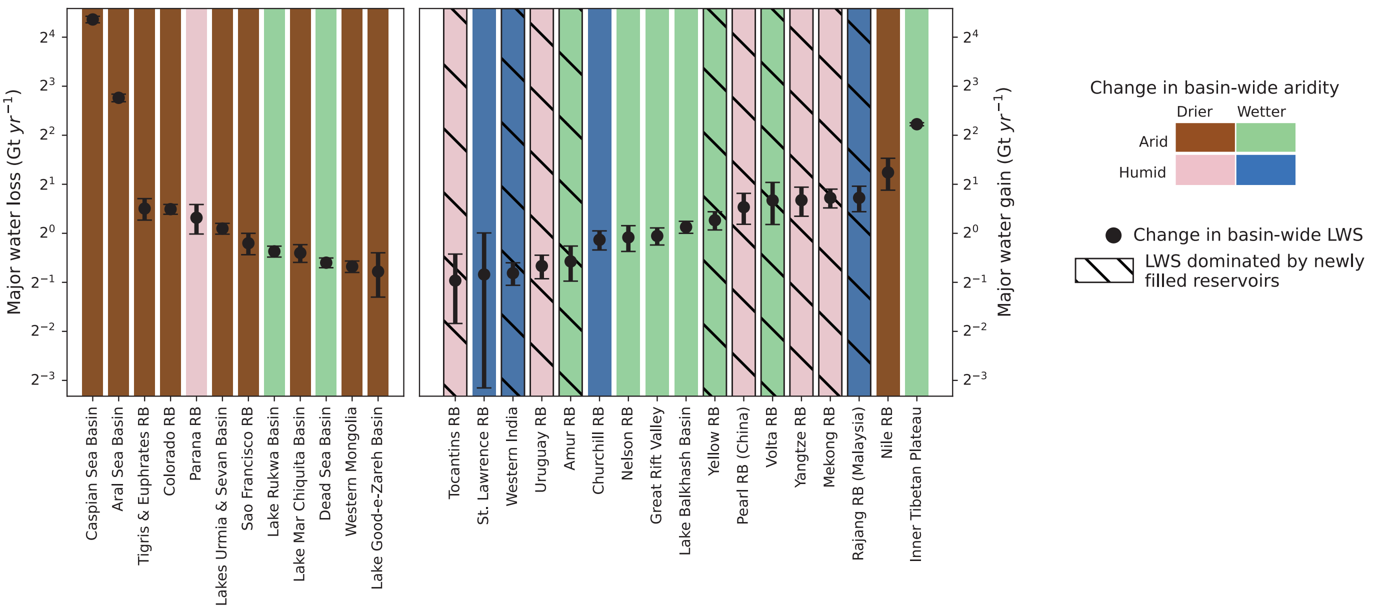


Fig. 4. Major losses and gains greater than 0.5 Gt year^{-1} in basin-wide lake water storage. Bar colors represent the trending direction in aridity in arid and humid basins. The hatched area indicates basins where the LWS trend was dominated by recently filled reservoirs. The error bars show the aggregate uncertainty in LWS trends at a 95% confidence interval. RB, for river basin. Major losses are primarily found in arid basins that are getting drier, and major gains in humid basins are mostly due to recent reservoir filling.

the newly filled reservoirs are concentrated in a few basins experiencing a dam construction boom, including the Nile, Yangtze, Mekong, Volta, Rajang, Pearl, Yellow, Amur, Uruguay, and Tocantins basins, as well as in western India (Fig. 4). About 55% of these regions experienced a drying climate during the study period, suggesting that human activities partially reversed the negative impact of climate on basin-wide surface water storage by impounding more water in reservoirs. Beyond these dam-construction hotbeds, most (65%) reservoirs experienced significant storage declines in humid regions (Fig. 1), indicating that the benefits of increased new reservoir storage are not evenly distributed. Most natural lakes were also in decline in humid regions, including high latitudes and the tropics (Fig. 1). However, these natural lake-volume losses were offset by precipitation-and-runoff–driven LWS gains in the northern Great Plains and the Laurentian Great Lakes of North America (Fig. 3B).

Discussion

We leveraged three decades of satellite observations to map long-term global LWS variability, finding evidence of widespread decline in global LWS with more than half of large water bodies (53 ± 2%) exhibiting significant drying trends. Roughly, one-quarter of the significant water losses in natural lakes were dominated by human activities or increasing temperature and PET, with the latter changes most often attributed to climate change (*41*, *42*). Before this study, many of the human and climate change footprints were either unknown, such as the desiccations of Lake Good-e-Zareh in Afghanistan and Lake Mar Chiquita in Argentina, or known only anecdotally, as in northern Eurasia and Canada. These water losses affect both the water and carbon cycles. For example, Arctic lakes are drying partially because of changes in temperature and PET (Fig. 3A), which is in line with broader climate changes toward increasing evaporative loss due to higher lake temperatures and reduced lake ice extents (*43*–*45*). Changes in runoff could also contain a climate change footprint through increasing evaporation (*46*, *47*). If we assume that some fraction of runoff-driven water losses are attributable to climate change, 43% of drying lakes were at least partially influenced by climate change as water losses were dominated by changes in temperature, PET, or runoff (fig. S2). Widespread LWS decline, particularly accompanied by rising lake temperatures, could reduce the amount of absorbed carbon dioxide and increase carbon emissions to the atmosphere given that lakes are hotspots of carbon cycling (*3*, *48*). Alternatively, climate change can increase LWS, particularly in the Tibetan Plateau, where glacier retreat and permafrost thawing partially drove alpine lake expansion (*8*, *40*), and in regions where precipitation is projected to increase under warming, such as the Upper Mississippi River Basin (*49*).

Our findings suggest that drying trends worldwide are more extensive than previously thought (*8*, *12*, *21*). Luo *et al.* (*21*) reported a net global increase in natural lake storage at a rate of 16.12 Gt year^{-1} and that most of the global lakes in humid regions gained water storage, whereas lakes in arid regions with high human water stress were generally in decline over 2003–2020. Although we confirm a "dry-get-drier" pattern in LWS, our findings also show widespread LWS decline in the humid tropics and high-latitude regions over the past three decades, as well as a net global decline (−26.38 ± 1.59 Gt year^{-1}) in natural lake storage. This contrast indicates that extrapolating the trends inferred from a brief time series could be uncertain, suggesting the necessity for long-term observations. For example, Luo *et al.* (*21*) concluded that Amazon lakes are largely expanding on the basis of LWS time series over two short periods (2003–2009 and 2018–2020). By contrast, the vast majority of large Amazon lakes in our study show a decreasing LWS trend because our detailed multidecadal LWS time series fully capture the impacts of major droughts during the early 2000s, 2010, and 2015. More broadly, our finding indicates that an intensified water cycle in a warming climate (*50*) may not result in increased water storage in humid regions, in part because of increasing land evapotranspiration (*51*) and potentially longer drought recovery times (*52*). The continued observations from ongoing and new satellites, in particular the Surface Water and Ocean Topography (SWOT) mission (recently launched in December 2022), will be useful to

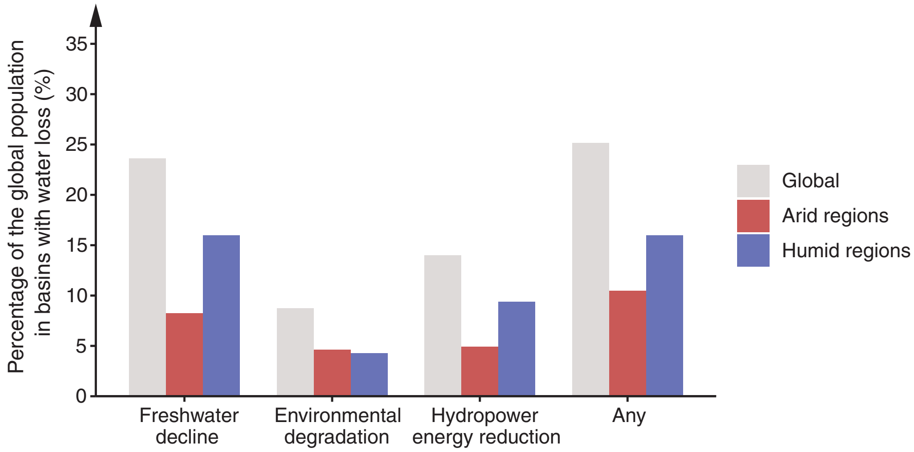


Fig. 5. Percentages of the global population residing in basins experiencing significant lake water loss. Affected sectors associated with lake water loss are freshwater decline, environmental degradation, and hydropower energy reduction. Approximately one-quarter of the global population resides in basins with lake water loss, partitioned comparably between arid and humid regions.

extend the LWS trends observed here and enable longer-term assessments of the interactions among LWS, the water cycle, and climate change.

This global-scale attribution of LWS trends has important implications for water resources management, particularly given that up to 2.0 billion people (one-quarter of the global population in 2023) reside in basins with large water bodies experiencing significant water storage losses (Fig. 5). Many of these drying lakes have been identified as important sources of water and energy (hydropower) (*22*) or listed among Ramsar sites of international importance (*53*). About 24, 9, and 14% of the global population resides in basins that are experiencing freshwater decline, environmental degradation, and hydropower energy reduction associated with decreasing LWS, respectively (Fig. 5). These population numbers can be roughly divided between arid and humid regions, with the latter having slightly larger shares. The reported populations are only estimates of potential impacts on lake basin residents who are likely the most affected by lake water loss (*4*, *54*). Sedimentation dominates the water loss in existing reservoirs, which suggests that reservoirs on which populations are heavily reliant will become less reliable for freshwater and hydroelectric energy supply because of aging. Sedimentation-induced storage water losses need to be considered in sustainable future freshwater supplies and long-term planning for reservoir operations and construction of new dams (*29*). More than half of the net loss in natural lake volume (56 ± 9%) is attributable to human activities and increasing temperature and PET, indicating that any recovery of water storage in these lakes could require substantial management efforts. Our findings are broadly consistent with existing studies (*7*, *27*, *55*) on human footprints on the Aral Sea, Dead Sea, and Lake Urmia but also reveal additional undocumented human-driven LWS losses in other large lakes, such as the Salton Sea in California and Lake Mar Chiquita in Argentina. The strongest attribution of human activities to LWS losses generally occurred in closed basins that were getting drier. This suggests that under conditions of declining precipitation, more-intensive human-water withdrawal from rivers led at least partially to the desiccations of closed lakes. Effective water conservation efforts can help save these water bodies (*4*), as shown in the success of recovering Lake Sevan through enacting water protection laws in Armenia (*28*). We detect that increasing temperature and PET are the chief determinants of water loss in 21% of drying natural lakes, a cautionary finding for a projected warmer future, underscoring the importance of accounting for climate change impacts within future surface water resources management.

REFERENCES AND NOTES

1. M. L. Messager, B. Lehner, G. Grill, I. Nedeva, O. Schmitt, *Nat. Commun.* **7**, 13603 (2016).
2. G. E. Hutchinson, *A Treatise on Limnology* (Wiley, ed. 1, 1957).
3. C. E. Williamson, J. E. Saros, W. F. Vincent, J. P. Smol, *Limnol. Oceanogr.* **54**, 2273–2282 (2009).
4. W. A. Wurtsbaugh *et al.*, *Nat. Geosci.* **10**, 816–821 (2017).
5. P. H. Gleick, *Int. Secur.* **18**, 79–112 (1993).
6. A. D. Gronewold, C. A. Stow, *Science* **343**, 1084–1085 (2014).
7. J.-F. Cretaux, R. Letolle, M. Bergé-Nguyen, *Global Planet. Change* **110**, 99–113 (2013).
8. J. Wang *et al.*, *Nat. Geosci.* **11**, 926–932 (2018).
9. L. Feng, C. Hu, X. Chen, X. Zhao, *Environ. Sci. Technol.* **47**, 9628–9634 (2013).
10. J.-F. Crétaux *et al.*, *Surv. Geophys.* **37**, 269–305 (2016).
11. J. L. Chen *et al.*, *Geophys. Res. Lett.* **44**, 6993–7001 (2017).
12. M. Rodell *et al.*, *Nature* **557**, 651–659 (2018).
13. J. Wang, Y. Sheng, Y. Wada, *Water Resour. Res.* **53**, 3854–3877 (2017).
14. J.-F. Pekel, A. Cottam, N. Gorelick, A. S. Belward, *Nature* **540**, 418–422 (2016).
15. Y. Wada *et al.*, *Hydrol. Earth Syst. Sci.* **21**, 4169–4193 (2017).
16. D. Tokuda, H. Kim, D. Yamazaki, T. Oki, *Geosci. Model Dev.* **14**, 5669–5693 (2021).
17. C. Schwatke, D. Dettmering, W. Bosch, F. Seitz, *Hydrol. Earth Syst. Sci.* **19**, 4345–4364 (2015).
18. T. Busker *et al.*, *Hydrol. Earth Syst. Sci.* **23**, 669–690 (2019).
19. H. Gao, C. Birkett, D. P. Lettenmaier, *Water Resour. Res.* **48**, W09504 (2012).
20. S. W. Cooley, J. C. Ryan, L. C. Smith, *Nature* **591**, 78–81 (2021).
21. S. Luo *et al.*, *Geophys. Res. Lett.* **49**, 1–10 (2022).
22. B. Lehner *et al.*, *Front. Ecol. Environ.* **9**, 494–502 (2011).
23. M. M. Tilzer, C. Serruya, *Large Lakes* (Springer Berlin Heidelberg, Berlin, Heidelberg, 1990).
24. F. Yao, J. Wang, C. Wang, J.-F. Crétaux, *Remote Sens. Environ.* **232**, 111210 (2019).
25. J. Wang *et al.*, *Earth Syst. Sci. Data* **14**, 1869–1899 (2022).
26. Y. Moon, U. Lall, H. Kwon, *Int. J. Climatol.* **28**, 361–370 (2007).
27. S. Chaudhari, F. Felfelani, S. Shin, Y. Pokhrel, *J. Hydrol. (Amst.)* **560**, 342–353 (2018).
28. A. Medvedev *et al.*, *Remote Sens. (Basel)* **12**, 3821 (2020).
29. T. J. Randle *et al.*, *J. Hydrol. (Amst.)* **602**, 126686 (2021).
30. W. L. Graf, E. Wohl, T. Sinha, J. L. Sabo, *Water Resour. Res.* **46**, W12535 (2010).
31. A. P. Williams *et al.*, *Science* **368**, 314–318 (2020).
32. A. Getirana, *J. Hydrometeorol.* **17**, 591–599 (2016).
33. N. Chao, Z. Luo, Z. Wang, T. Jin, *Ground Water* **56**, 770–782 (2018).
34. V. Mishra, K. Thirumalai, S. Jain, S. Aadhar, *Environ. Res. Lett.* **16**, 054007 (2021).
35. M. Venkatappa, N. Sasaki, P. Han, I. Abe, *Sci. Total Environ.* **795**, 148829 (2021).
36. A. I. J. M. van Dijk *et al.*, *Water Resour. Res.* **49**, 1040–1057 (2013).
37. T. A. Brás, J. Seixas, N. Carvalhais, J. Jägermeyr, *Environ. Res. Lett.* **16**, 065012 (2021).
38. M. Xiao, B. Udall, D. P. Lettenmaier, *Water Resour. Res.* **54**, 6739–6756 (2018).
39. M. Khaki, J. Awange, *Sensors* **21**, 4304 (2021).
40. F. Yao *et al.*, *Environ. Res. Lett.* **13**, 064011 (2018).
41. R. A. Kerr, *Science* **316**, 188–190 (2007).
42. D. G. Kingston, M. C. Todd, R. G. Taylor, J. R. Thompson, N. W. Arnell, *Geophys. Res. Lett.* **36**, L20403 (2009).
43. C. M. O'Reilly *et al.*, *Geophys. Res. Lett.* **42**, 10773–10781 (2015).
44. S. Sharma *et al.*, *Nat. Clim. Chang.* **9**, 227–231 (2019).
45. G. Zhao, Y. Li, L. Zhou, H. Gao, *Nat. Commun.* **13**, 3686 (2022).
46. N. S. Christensen, A. W. Wood, N. Voisin, D. P. Lettenmaier, R. N. Palmer, *Clim. Change* **62**, 337–363 (2004).
47. J. R. Miller, G. L. Russell, *J. Geophys. Res.* **97** (D3), 2757 (1992).
48. J. A. Rosentreter *et al.*, *Nat. Geosci.* **14**, 225–230 (2021).
49. E. Sinha, A. M. Michalak, V. Balaji, *Science* **357**, 405–408 (2017).
50. M. R. Allen, W. J. Ingram, *Nature* **419**, 224–232 (2002).
51. D. G. Miralles *et al.*, *Nat. Clim. Chang.* **4**, 122–126 (2014).
52. C. R. Schwalm *et al.*, *Nature* **548**, 202–205 (2017).
53. *Ramsar Sites Information Service* (Ramsar, accessed 24 November 2021); https://rsis.ramsar.org/.
54. W. S. Breffle *et al.*, *Resour. Policy* **38**, 152–161 (2013).
55. R. A. Al-Weshah, *Hydrol. Process.* **14**, 145–154 (2000).

ACKNOWLEDGMENTS

We thank A. Temme, C. Martin, S. Nerem, and J. Rosentreter for comments on the results and K. Yang and K. Bogan for suggestions and assistance on the visualizations. We thank P. Lin for sharing the latest version of the GRADES data and H. Beck for sharing the latest version of the MSWEP data for our analyses. Assistance for processing climate and human water use datasets was provided by K. Yang. We also thank L. Patterson for help with downloading the in situ data of reservoirs and L. Pitcher and E. Knight for proofreading our manuscript. We thank C. Schwatke, C. Birkett, and S. Cooley for making the altimetry-derived water level data publicly available. **Funding:** This study was supported by NOAA Cooperative Agreement with CIRES (NA17OAR4320101) to F.Y., NASA NIP grant (80NSSC18K0951) to B.L., and the "Climate Change Initiative Grant (4000125030/18/I-NB) to J.-F.C. and M.B.-N." **Author contributions:** F.Y. conceptualized the project. F.Y., J.W., J.-F.C., and M.B.-N. estimated and analyzed

the water storage variability and trends. F.Y., B.R., and B.L. developed the statistical models. F.Y. and B.R. estimated the sedimentation-induced storage loss in reservoirs. F.Y. evaluated the potential impacts of drying lakes on population with inputs from B.L., B.R., and Y.W. F.Y. conducted the validation. F.Y. performed the visualization with inputs from B.L., J.W., and Y.W. F.Y. wrote the manuscript with inputs from B.L., B.R., and J.W. All authors read and commented on drafts of this paper. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The Landsat images, including Landsat 5 Thematic Mapper, Landsat 7 Enhanced Thematic Mapper-plus, and Landsat 8 Operational Land Imager, are available from the US Geological Survey at http://earthexplorer.usgs.gov and the Google Earth Engine platform at https://earthengine.google.com. ICESat and ICESat-2 data are available from the National Snow and Ice Data Center (NSIDC) at https://nsidc.org/data. Water levels derived from ICESat-2 are available at https://doi.org/10.5281/zenodo.4489056. Water level products from radar altimeters can be downloaded from the Hydroweb at http://hydroweb.theia-land.fr, the Database for Hydrological Time Series of Inland Waters (DAHITI) at https://dahiti.dgfi.tum.de/en, and the USDA Global Reservoir and Lake Monitor database at https://ipad.fas.usda.gov/cropexplorer/global_reservoir. The CryoSat-2 data are available from the European Space Agency (ESA) at https://earth.esa.int/eogateway/catalog/cryosat-products. The Global Reservoir Bathymetry Dataset can be downloaded from https://dataverse.tdl.org/dataset.xhtml?persistentId=doi:10.18738/T8/TO5HJG. The Global Surface Water (GSW) dataset is available from https://global-surface-water.appspot.com/ and the Google Earth Engine platform at https://earthengine.google.com. Reservoir sedimentation survey data from the US Army Corps can be accessed at https://water.usace.army.mil/ and https://nicholasinstitute.duke.edu/reservoir-national-trends/sediment/. USGS gauge data can be downloaded from https://waterdata.usgs.gov/nwis/, US Army Corps gauge data can be downloaded from https://water.usace.army.mil/ and https://nicholasinstitute.duke.edu/reservoir-data/, California Department of Water Resources gauge data can be downloaded from https://cdec.water.ca.gov/, gauge data from Texas Water Development Board can be downloaded from https://waterdatafortexas.org/reservoirs/statewide, gauge data from Spain can be downloaded from https://ceh.cedex.es/anuarioaforos/afo/embalse-nombre.asp, and gauge data from Bureau of Meteorology in Australia can be downloaded from http://www.bom.gov.au/waterdata/. The HydroLAKES database can be downloaded from https://www.hydrosheds.org/page/hydrolakes. The Georeferenced global Dams And Reservoir dataset (GeoDAR) can be downloaded from https://doi.org/10.5281/zenodo.6163413. The database of Roller-Compacted Concrete (RCC) dams can be accessed at http://www.rccdams.co.uk/. The Global Lake area, Climate, and Population (GLCP) dataset can be downloaded at https://portal.edirepository.org/nis/mapbrowse?packageid=edi.394.4. The HydroSHEDS dataset can be downloaded from https://hydrosheds.org/page/overview. The Climatic Research Unit (CRU) data are available from https://crudata.uea.ac.uk/cru/data/hrg/. ECMWF Reanalysis v5 (ERA5) data are available from https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5. Modern-Era Retrospective analysis for Research and Applications version 2 (MERRA-2) are available from https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/. Multi-Source Weighted-Ensemble Precipitation (MSWEP) are available from https://www.gloh2o.org/mswep/. Global Historical Climatology Network data are available from https://www.ncei.noaa.gov/products/land-based-station/global-historical-climatology-network-monthly. Global Land Evaporation Amsterdam Model (GLEAM) data are available from https://www.gleam.eu/. Global Reach-scale A priori Discharge Estimates for SWOT (GRADES) dataset can be downloaded from http://hydrology.princeton.edu/data/mpan/GRFR/discharge/daily/. The reconstructed human water use data derived from four global hydrologic models can be downloaded from https://zenodo.org/record/1209296#.YZPcr2DMKM8. The water body masks delineated from the GSW dataset, lake volume time series derived from Landsat images and satellite altimeters, lake volume trend estimates, attribution, and all validation analyses are available on the Zenodo data repository at https://zenodo.org/record/7946043. Water storage trends and drivers of studied large lakes are available in the interactive map at https://cires.colorado.edu/globallakes. **Code availability:** R scripts that were used to process hydroclimate and human water use data, to derive water levels from ICESat and ICESat-2, to construct lake water storage time series, to estimate the mean rate of sedimentation in reservoirs, to estimate trends in lake water storage and basin aridity, to conduct validation, to construct regression model ensemble, and to estimate affected population are available on CodeOcean at https://codeocean.com/capsule/0322198/tree/v1. JavaScript scripts for mapping water areas from Landsat images and IDL Scripts for processing CyoSat-2 data are available in the supplementary materials. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abo2812
Materials and Methods
Figs. S1 to S11
References (*56*–*100*)

Submitted 11 March 2022; accepted 31 March 2023
10.1126/science.abo2812

CARBON SEQUESTRATION

Releasing global forests from human management: How much more carbon could be stored?

Caspar T. J. Roebroek[1,2]*, Gregory Duveiller[3], Sonia I. Seneviratne[2], Edouard L. Davin[4,5,6], Alessandro Cescatti[1]*

Carbon storage in forests is a cornerstone of policy-making to prevent global warming from exceeding 1.5°C. However, the global impact of management (for example, harvesting) on the carbon budget of forests remains poorly quantified. We integrated global maps of forest biomass and management with machine learning to show that by removing human intervention, under current climatic conditions and carbon dioxide (CO_2) concentration, existing global forests could increase their aboveground biomass by up to 44.1 (error range: 21.0 to 63.0) petagrams of carbon. This is an increase of 15 to 16% over current levels, equating to about 4 years of current anthropogenic CO_2 emissions. Therefore, without strong reductions in emissions, this strategy holds low mitigation potential, and the forest sink should be preserved to offset residual carbon emissions rather than to compensate for present emissions levels.

Forests play a key role in the global carbon cycle (*1*) by capturing and storing carbon in tree biomass and enhancing soil organic matter. As such, they are a key component in global policy-making to mitigate climate change. Most of the main strategies aiming to keep global warming below a threshold of 1.5°C rely on forest-based mitigation plans as approaches to complement hard reductions in anthropogenic carbon emissions to reach carbon neutrality (*2*). In these scenarios, forests constitute one of the main carbon sinks that offset continuing emissions (*3*) from agriculture and other sectors that will require a longer time frame to decarbonize or are unlikely to reach effective net-zero emissions (*4*).

In past decades, intact forests have provided a strong carbon sink (*5*) thanks to the fertilization effect of the increasing CO_2 concentration on photosynthesis, which has spurred increased tree growth and carbon storage. However, the CO_2-fertilization effect, although still positive, is likely in decline because of the increasing importance of other limiting factors (*6*). Consequentially, the terrestrial carbon sink will likely stabilize (*3*). At the same time, climate change is increasingly pushing tree populations to the edge of their natural climate envelopes (*7*, *8*) and increasing the frequency and intensity of natural disturbances, leading to higher risks of tree mortality (*9*, *10*). This situation may change forests that are now seen as reliable carbon sinks into active net carbon sources. These processes shift the perspective of forest-based climate change mitigation toward two strategies: adapting management practices in forests that are currently below their natural equilibrium biomass level (*11*) and expanding forest areas through afforestation and reforestation programs.

Both strategies present substantial complications in their implementation, as documented in the Intergovernmental Panel on Climate Change (IPCC) special report on land (*4*). Afforestation and reforestation potential is mainly limited by the strong competition for land (*12*). Adapting management practices in forests—for example, reducing wood harvesting—has the potential to increase the carbon stock (*11*, *13*–*15*). However, this potential is limited by the societal demand for wood and other forest-based products, which also provide climate benefits by substituting for energy and materials that would have contributed more severely to climate change, such as fossil fuels and concrete. Nevertheless, recent research shows that reducing harvesting intensity has strong climate mitigation benefits, especially at shorter timescales, even after accounting for such substitutions (*15*). However, on the long term and at a global scale, the total amount of carbon that could realistically be stored in forests remains poorly understood (*16*).

In this study, we focused on the natural limits to additional carbon accumulation in the biomass of existing forests, aiming to quantify how much more carbon could realistically be stored in the hypothetical scenario in which all current forests reach natural equilibrium. This hypothetical scenario defines the upper bound of carbon storage in

[1]European Commission, Joint Research Centre (JRC), Ispra, Italy. [2]Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland. [3]Max Planck Institute for Biogeochemistry, Jena, Germany. [4]Wyss Academy for Nature, University of Bern, Bern, Switzerland. [5]Climate and Environmental Physics, Physics Institute, University of Bern, Bern, Switzerland. [6]Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland.
*Corresponding author. Email: caspar.roebroek@ec.europa.eu (C.T.J.R.); alessandro.cescatti@ec.europa.eu (A.C.)

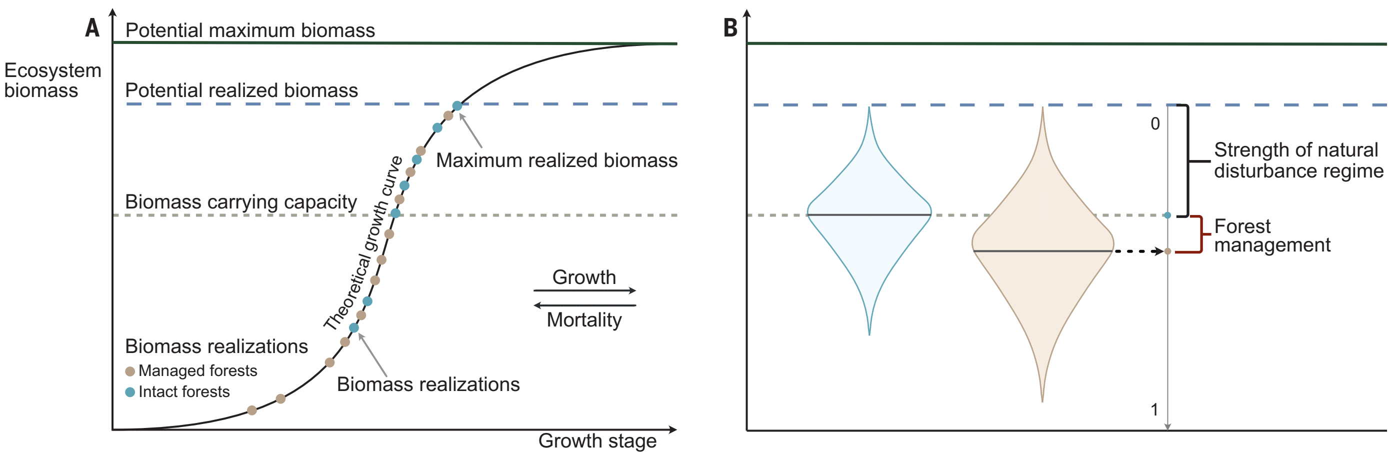


Fig. 1. Graphical representation of both the natural and human-altered forest biomass budget and the impact of natural and anthropogenic disturbance regimes. (**A**) Idealized temporal dynamics of biomass in an undisturbed system (curved black line) that translates across the landscape in a series of realized forest patches (blue and brown dots) at various biomass levels as a result of natural and anthropogenic disturbances. (**B**) When aggregated at a regional scale, biomass levels in forest patches can be described with frequency distributions (violin plots) that will differ between managed and intact forests because of direct human management (brown bracket). This biomass gap is finally interpreted as the potential for additional carbon storage of forests if human intervention were to be suspended.

existing forests and implies the removal of all direct human management from forests (meaning any human-caused physical changes in forest structure, carbon storage, or species composition that are due to harvesting, fire suppression, plantation, or other factors). Throughout this text, we refer to forests that have been altered by direct human action simply as "managed forests" as opposed to "intact forests." Some human interventions, such as active fire suppression, could in practice enhance carbon stored in forests, but from a global perspective, managed forests contain substantially less carbon than their intact counterparts (*11, 16*). We focused on existing forests because other studies suggest that they hold most of the additional carbon storage potential of forests, considering management as well as afforestation and reforestation when accounting for competition with other land uses (*16*).

Defining additional carbon storage potential

The additional carbon storage potential of forests has been studied indirectly as part of estimating carbon storage potential on land, including afforestation and reforestation as well as removal of all human management from forests (*11*, *16*, *17*). However, previous studies did not strictly account for and separate the effect of natural disturbances and human interventions on potential carbon storage in forests. They instead targeted potential maximum biomass in the absence of both natural disturbances and management (*16*), thus creating an upper bound estimate, which likely overestimates the realistic residual carbon storage potential of forests (*17*).

We approached the realistic potential additional carbon storage in forests as the difference between the existing carbon stock and the current carbon carrying capacity. Carbon carrying capacity is defined as the carbon "stored in a forest ecosystem under prevailing environmental conditions and natural disturbance regimes, but excluding anthropogenic disturbance" (*18*) (Fig. 1), in which anthropogenic disturbances can be understood as any direct human management in forests. Carbon carrying capacity is an inherent trait of a climate state and thus excludes any future changes in forest growth potential, likelihood of natural disturbances, and enhanced growth due to higher atmospheric CO_2 concentrations. In the context of biomass, we refer to it as "biomass carrying capacity," and it is directly translatable with the general rule of thumb that biomass contains 50% carbon.

We propose a modeling framework that calculates the biomass carrying capacity and expected biomass occurring under local conditions—including both natural disturbances and direct human management (which we refer to as "expected biomass")—in two steps [Fig. 1 and (*19*)]. First, potential maximum biomass for each pixel is modeled from biomass observations [GlobBiomass, which reproduces country-aggregated field observations with a coefficient of determination (R^2) of 0.96 (*20*)]. Second, disturbance regimes are calculated as the average ratio between current biomass and the potential biomass [1 – (biomass/potential biomass)]. Forests persisting close to the potential are thus attributed a low disturbance-regime value and vice versa. This is done twice: The natural disturbance regime is derived from the ratio found in intact forests, whereas the total (natural disturbance regime and management) uses the ratio found in all forests (intact and managed). Multiplying the potential maximum biomass by the natural disturbance–regime ratio yields the biomass carrying capacity, whereas multiplying by the total disturbance regime results in the expected biomass. Furthermore, the difference between biomass carrying capacity and expected biomass is interpreted as the net effect of all direct human management, a measure of the

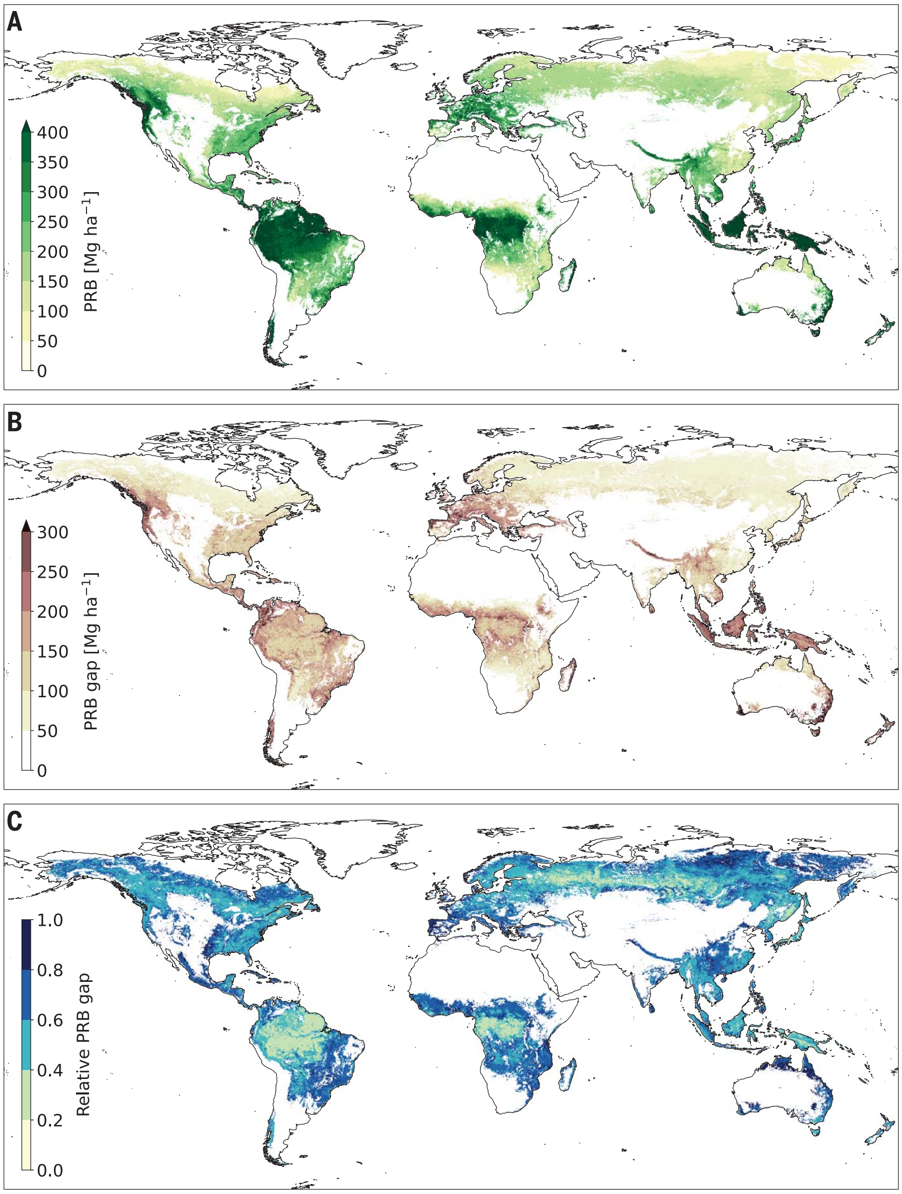


Fig. 2. Global representation of the potential realized biomass and its comparison with observed biomass values. Maps of (**A**) the absolute values of potential realized biomass (PRB), the maximum aboveground biomass that could occur without natural disturbances, (**B**) the absolute difference between PRB and the currently observed biomass, and (**C**) the relative difference between biomass observations and PRB [1 – (observed biomass/potential realized biomass)].

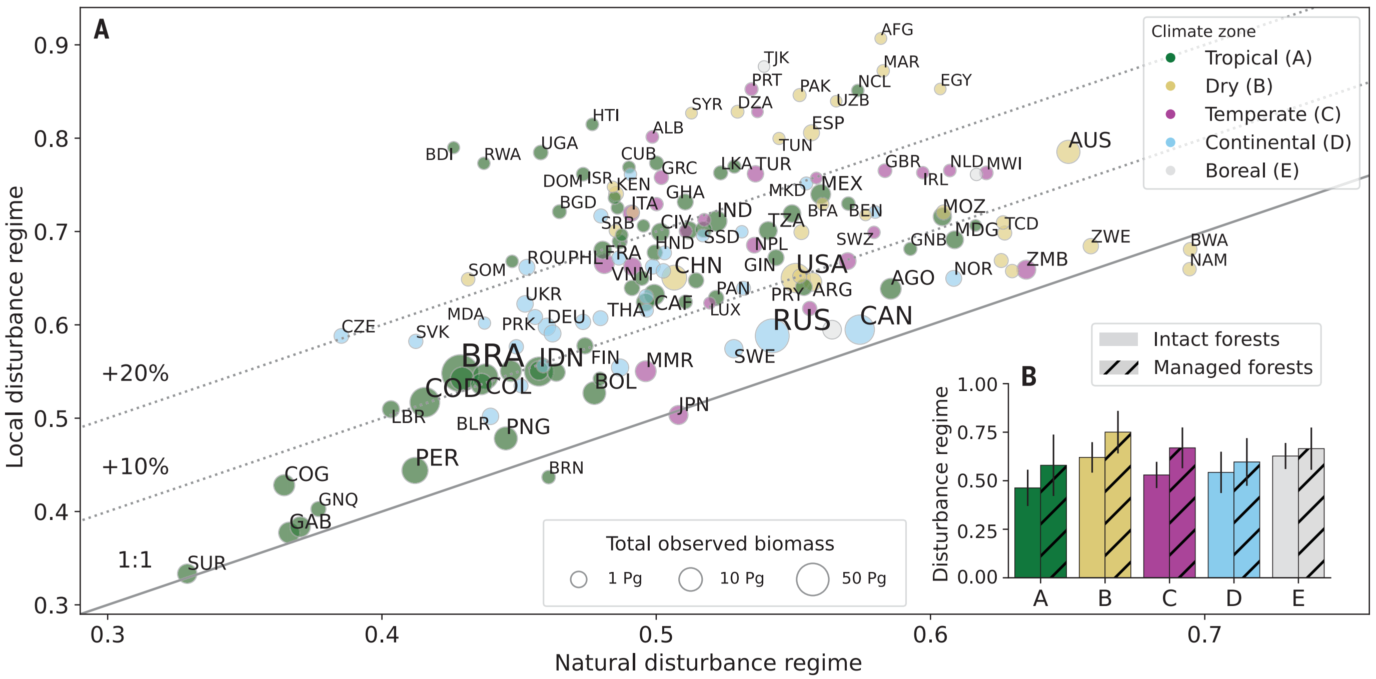


Fig. 3. Difference between natural and total disturbance regimes. (**A**) Countries are shown with their natural and total (natural plus direct human management) dsturbance regimes. Most countries display a substantial increase in the disturbance regime when current disturbance regimes are compared with those expected in equivalent intact forests. Within-country variability is large, with the highest increases in the impact of disturbance regimes almost tripling. Country abbreviations are defined at https://www.iso.org/obp/ui/#search/code/. (**B**) The same results noted in (A) are shown per the Köppen-Geiger climate zones. All climate zones see a substantial increase in disturbance regime when accounting for human interventions. Disturbance regimes are reported as fractions of potential realized biomass (Fig. 1).

additional carbon storage potential from theoretically removing all human management.

Global variation in potential realized biomass

The first step in the framework is to model potential maximum biomass. Because the values are derived from biomass observations and because the true potential will not always be present in the data, we refer to our model estimates as "potential realized biomass" (Fig. 1) (*21*–*23*). We chose a quantile machine-learning method [quantile random forest (*24*); see (*19*) and figs. S1 and S2] to model potential realized biomass because of its ability to deal with high-dimensional and nonlinear relationships (*24*). We modeled forest biomass (from GlobBiomass) as a function of the environmental characteristics (climate features, physical soil characteristics, and soil water availability; table S1), with the assumption that forests in similar conditions have similar potential realized biomass. A comparable approach was used by Walker *et al.* (*16*), with the difference being that we used a quantile regression approach (calculating the maxima rather than the mean) to focus strictly on potentials.

We found the highest values of potential realized biomass, around 600 Mg biomass/ha, to be in both the humid tropics and humid temperate zones, whereas boreal forests may contain up to 150 Mg biomass/ha (Fig. 2A). Tropical forests persist close to their maximum potential biomass (Figs. 2B and 2C), although a minimum difference of 10 to 15% between potential realized biomass and observed biomass is still to be expected in most of the tropics because of natural age dynamics and a relatively low frequency of natural disturbances (*25*). The biomass of temperate forests can also reach very large values yet experience stronger natural disturbance regimes, with forest fires, windthrows, and insect outbreaks being an integral part of the ecosystem dynamics (*26*). This difference is reflected in greater spatial variation in the original biomass data, which translates to larger differences between potential and observed forest biomass. Europe in particular shows a large difference between potential and realized biomass levels as a result of the long history of land management.

Effects of disturbance on biomass

The second step in the modeling framework is to calculate the disturbance regimes. The natural disturbance regimes were modeled similarly to potential realized biomass, as a function of environmental variables, assuming that how close an intact forest is on average to the potential realized biomass is again driven by the environmental context of the forest. In contrast to the potential realized biomass model, we used a standard (mean) random forest regression model, trained on forests that fall within intact forest landscapes (*27*). The total disturbance regimes (natural plus direct human management) were calculated by means of a similar random forest machine-learning model but with geographical coordinates for predictions instead of climate gradients. A comparison between the natural and the total, human-affected disturbance regimes (Fig. 3) shows that in most countries, the effective total disturbance regime has at least a 10% greater impact on biomass than what would be expected from natural disturbances alone. In the 10 countries with the highest total potential realized biomass, an increase of 20% is observed as compared with the natural level. Locally, the direct impacts of human interventions are greater, such as in some areas of Eastern South America, Western Africa, Europe, and China, where we show more than double the disturbance impacts in managed forests when compared with the equivalent intact forests.

Biomass carrying capacity of intact forests

By multiplying the previously obtained natural disturbance–regime ratio (fig. S8) by the potential realized biomass, we calculated the biomass carrying capacity—that is, the biomass that can realistically be expected to be obtained in intact forests given the natural disturbance regime [Eq. 2 in (*19*)] (fig. S9). In

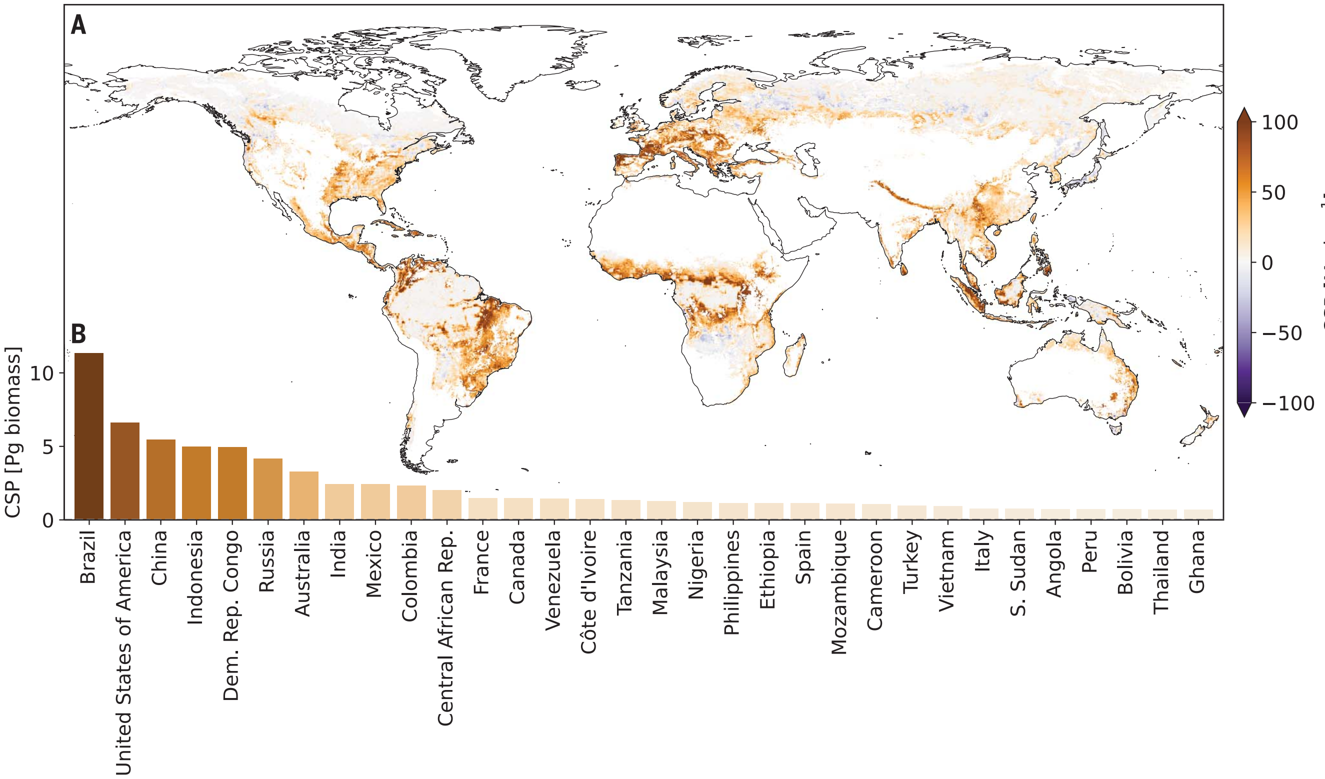


Fig. 4. Additional carbon storage potential. (**A**) Additional carbon storage potential (CSP) in the hypothetical scenario in which all forests would resettle in their natural equilibrium if all direct human management was removed from them. The CSP is calculated from the difference between biomass carrying capacity and the expected biomass (the biomass that would occur under local conditions with the given natural disturbance regime and average intensity of human intervention). (**B**) National statistics of additional CSP for countries where absolute values exceed 0.7 Pg biomass.

comparison with the average biomass per ecosystem type used in other studies (*5*, *17*), our average biomass carrying capacity values are slightly smaller—for example, the average in natural tropical rainforests is approximately 8% less at 260 Mg biomass/ha (*5*), although some areas surpass 390 Mg biomass/ha. This difference is likely due to the careful accounting of the impact of climate-driven natural disturbances on the forest biomass budget that can be achieved with our methodology and to a lower bias resulting from a more representative sample of forests per ecosystem type.

Additional carbon storage potential

Additional carbon storage potential, which is the direct opposite of the net effect of direct human management in forests, was calculated as the difference between the biomass carrying capacity and expected biomass. Almost all world regions show additional carbon storage potential (Fig. 4), with the largest potentials concentrated in the regions on the edges of tropical rainforests in South America and Africa, corresponding to areas that have experienced high rates of deforestation and forest degradation (*28*). Intact ecosystems (fig. S10), identified from remote sensing data as regions with no sign of human intervention (*27*), show close-to-zero differences between the natural and total disturbance regimes (globally average deviation biomass carrying capacity and observed biomass is 0.3 Mg biomass/ha).

The boreal regions show limited potential, with some areas even displaying negative carbon storage potential. These negative values might be explained by several hypotheses: (i) In these regions, intact forests are not fully comparable to the managed forests because of the preferred selection of the most fertile sites for forest management. (ii) Management increased the maximum potential realized biomass (for example, through the introduction of highly productive monospecific stands) or deliberately decreased the natural disturbance regime (for example, active and passive suppression of forest fires). (iii) Small differences appeared, with a relatively large effect on biomass potentials, in environmental conditions that are not resolved in the input data.

Our analysis estimates a total additional carbon storage potential in current forests of 88.1 Pg [error range: 42.0 to 125.9 Pg biomass; for error range definition, see (*19*)] of aboveground biomass (AGB) (15.2 to 16.1% more than the biomass reported in current forests, with large differences between countries) following the hypothetical scenario of no human intervention, which is roughly equivalent to 44.1 Pg of carbon (PgC) (error range: 21.0 to 63.0 PgC). About 37% of the carbon storage potential is located in five countries: Brazil, United States, China, Democratic Republic of the Congo, and Indonesia. This figure is less than half of the values found in previous work [for example, Walker *et al.*: 111.6 PgC AGB (*16*)], likely as a result of our strict accounting of naturally occurring disturbances. By comparison, recent work found that carbon storage potential of afforestation and reforestation is around 15.6 PgC of AGB when accounting for land-related constraints (*16*).

Conclusions

The modeling results were validated with a series of uncertainty analyses (*19*) (figs. S3 to S6), and estimates of the biomass carrying capacity were compared with the biomass observations of the intact forests (R^2 = 0.99; fig. S7A). We found that relative uncertainties of the biomass carrying capacity are approximately equal to the original biomass data (fig. S3). The largest uncertainties are found in the temperate regions, where few intact forests are available for the extrapolation of the disturbance regime. The effect of these uncertainties on the global values of additional carbon storage potential is relatively minor [(*19*), uncertainty analysis 5]. The

results allow us to draw some highly relevant conclusions on the overall potential magnitude of forest-based mitigation options on the basis of the increase of carbon stocks.

First, carbon that can additionally be stored in the global aboveground forest biomass corresponds to about four times the amount of total human emissions released in 2019 (around 10 PgC). Although this is a substantial amount and an underestimation of the full potential because we only considered AGB, it corresponds to the emissions within a shorter period than the time that has elapsed since the Paris Agreement of 2015 and would require the full cessation of forest management. Second, all forests together now contain about four times as much carbon as could additionally be stored in their AGB (GlobBiomass values). This carbon is already at risk of being partially released into the atmosphere by forest degradation (*27*, *29*, *30*) and increased natural disturbances (*7*, *8*, *10*). Third, obtaining the equivalent of the potential increase of carbon in currently managed forests through reforestation and afforestation would require a land area of about 7.1 million km^2 (an area almost as large as Australia) (using an average biomass carrying capacity of 123 Mg biomass/ha). To dedicate enough land to achieve this number would be a massive undertaking, and furthermore, this carbon stock would require decades to fully develop.

These observations lead to the conclusion that current forests have only a limited additional carbon storage potential to substantially mitigate the increase in atmospheric CO_2 without major reductions to fossil emissions. For this reason, forest-based climate mitigation should not be mistaken as a trade-off for present carbon emissions levels. On the contrary, efforts to reverse trends in deforestation and increase carbon stocks should be seen as precious and effective strategies to offset future residual emissions from the agricultural sector and from essential industries that are unlikely to reach net-zero emissions.

REFERENCES AND NOTES

1. G. B. Bonan, *Science* **320**, 1444–1449 (2008).
2. O. Hoegh-Guldberg *et al.*, in *Global Warming of 1.5°C* (IPCC, 2018), p. 138.
3. P. Friedlingstein *et al.*, *Earth Syst. Sci. Data* **14**, 4811–4900 (2022).
4. G. Jia *et al.*, in *Climate Change and Land* (IPCC, 2019), p. 118.
5. Y. Pan *et al.*, *Science* **333**, 988–993 (2011).
6. S. Wang *et al.*, *Science* **370**, 1295–1300 (2020).
7. A. Mauri *et al.*, *Sci. Data* **9**, 37 (2022).
8. C. A. Boulton, T. M. Lenton, N. Boers, *Nat. Clim. Change* **12**, 271–278 (2022).
9. G. Forzieri *et al.*, *Nat. Commun.* **12**, 1081 (2021).
10. D. Bauman *et al.*, *Nature* **608**, 528–533 (2022).
11. K.-H. Erb *et al.*, *Nature* **553**, 73–76 (2018).
12. Z. Sha *et al.*, *Commun. Earth Environ.* **3**, 8 (2022).
13. S. Luyssaert *et al.*, *Nature* **562**, 259–262 (2018).
14. W. R. Moomaw, B. E. Law, S. J. Goetz, *Environ. Res. Lett.* **15**, 045009 (2020).
15. T. Skytt, G. Englund, B.-G. Jonsson, *Environ. Res. Lett.* **16**, 114037 (2021).
16. W. S. Walker *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2111312119 (2022).
17. J.-F. Bastin *et al.*, *Science* **365**, 76–79 (2019).
18. H. Keith, B. G. Mackey, D. B. Lindenmayer, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 11635–11640 (2009).
19. Materials and methods are available as supplementary materials.
20. M. Santoro, GlobBiomass - global datasets of forest biomass, PANGAEA (2018); https://doi.org/10.1594/PANGAEA.894711.
21. W. H. McNab, *For. Sci.* **35**, 91–104 (1989).
22. H. Y. Chen, K. Klinka, R. D. Kabzems, *Can. J. For. Res.* **28**, 1743–1755 (1998).
23. T. Scholten *et al.*, *J. Plant Ecol.* **10**, 111–127 (2017).
24. N. Meinshausen, *J. Mach. Learn. Res.* **7**, 983–999 (2006).
25. H. C. Muller-Landau *et al.*, *New Phytol.* **229**, 3065–3087 (2021).
26. P. van Lierop, E. Lindquist, S. Sathyapala, G. Franceschini, *For. Ecol. Manage.* **352**, 78–88 (2015).
27. P. Potapov *et al.*, *Sci. Adv.* **3**, e1600821 (2017).
28. C. Vancutsem *et al.*, *Sci. Adv.* **7**, eabe1603 (2021).
29. Y. Qin *et al.*, *Nat. Clim. Change* **11**, 442–448 (2021).
30. M. C. Hansen *et al.*, *Science* **342**, 850–853 (2013).
31. M. Buchhorn *et al.*, Copernicus Global Land Service: Land Cover 100m: version 3 Globe 2015-2019: Algorithm Theoretical Basis Document, Zenodo (2020); https://doi.org/10.5281/zenodo.3938968.
32. S. E. Fick, R. J. Hijmans, *Int. J. Climatol.* **37**, 4302–4315 (2017).
33. L. Poggio *et al.*, *Soil* **7**, 217–240 (2021).
34. J. Muñoz Sabater, ERA5-Land hourly data from 1950 to present (2019); https://doi.org/10.24381/cds.e2161bac.
35. Y. Fan, G. Miguez-Macho, E. G. Jobbágy, R. B. Jackson, C. Otero-Casal, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 10572–10577 (2017).
36. J. J. Danielson, D. B. Gesch, "Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010)," (US Geological Survey, 2011); https://pubs.usgs.gov/of/2011/1073/pdf/of2011-1073.pdf.
37. C. T. J. Roebroek, jasperroebroek/biomass_carrying_capacity: publication, Zenodo (2022); https://doi.org/10.5281/zenodo.7486354.
38. C. T. J. Roebroek, Supplement to: "Releasing global forests from management: How much more carbon could be stored?" Figshare (2022); https://doi.org/10.6084/m9.figshare.21781916.v1.

ACKNOWLEDGMENTS

Funding: This work was realized with the collaboration of the European Commission Joint Research Center and ETH Zürich, under Collaborative Doctoral Partnership Agreement no. 35317. **Author contributions:** Conceptualization was done by all coauthors. C.T.J.R. designed the methodology with the contribution of A.C. C.T.J.R. performed the formal analysis. The text writing of the original draft was done by C.T.J.R., A.C., G.D., and E.L.D., and S.I.S. supervised the work. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The datasets used for the analysis presented in this manuscript are listed in table S1 and are publicly available. We used the GlobBiomass dataset (*20*), Copernicus Global Land Service: Land Cover 100 version 3 landcover data (*31*), bioclimatic variables from WorldClim version 2.1 (*32*), physical soil characteristics from SoilGrids version 2 (*33*), soil moisture data from ERA5-Land (*34*), water-table depth data from Fan *et al.* (*35*), and the GMTED2010 digital elevation model (*36*). The code (*37*) and produced results (*38*) of this study are both available online.

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.add5878
Materials and Methods
Figs. S1 to S10
Table S1
References (*39–41*)
MDAR Reproducibility Checklist

Submitted 21 June 2022; accepted 14 April 2023
10.1126/science.add5878

BEHAVIORAL ECOLOGY

Fear of large carnivores amplifies human-caused mortality for mesopredators

Laura R. Prugh[1]*, Calum X. Cunningham[1], Rebecca M. Windell[1], Brian N. Kertson[2], Taylor R. Ganz[1], Savanah L. Walker[3], Aaron J. Wirsing[1]

The challenge that large carnivores face in coexisting with humans calls into question their ability to carry out critical ecosystem functions such as mesopredator suppression outside protected areas. In this study, we examined the movements and fates of mesopredators and large carnivores across rural landscapes characterized by substantial human influences. Mesopredators shifted their movements toward areas with twofold-greater human influence in regions occupied by large carnivores, indicating that they perceived humans to be less of a threat. However, rather than shielding mesopredators, human-caused mortality was more than three times higher than large carnivore–caused mortality. Mesopredator suppression by apex predators may thus be amplified, rather than dampened, outside protected areas, because fear of large carnivores drives mesopredators into areas of even greater risk from human super predators.

Apex predators can exert a strong influence on ecosystems through cascading top-down effects on lower trophic levels (*1*). Human persecution and habitat loss have reduced the ranges of large carnivores by an average of ~50% worldwide (*2*), and most large-mammal populations persist outside protected areas where they remain subject to strong human influence (*3*). Persecution outside protected areas can even reduce populations within protected areas because of source-sink dynamics (*4*). Sustaining functional populations of wide-ranging species such as carnivores thus requires understanding dynamics outside protected areas (*5*). However, the ecological roles of large carnivores have been investigated largely within protected areas (*6*), which leaves a critical gap in understanding the ecosystem functioning of human-dominated landscapes.

Even within protected areas, human presence can substantially alter wildlife behavior, with large carnivores often avoiding areas frequented by human visitors (*7*). This avoidance can lead to increased use of human-affected areas by subordinate species, a phenomenon known as the "human shield" effect (*8–10*). The use of human shields in areas where hunting by people is restricted can indeed yield increased survival rates (*11*, *12*). Outside protected areas, however, human shields can become lethal for ungulates and mesopredators that may preferentially use anthropogenic areas to avoid large carnivores. Human shields can impose mortality risks on wildlife through hunting and trapping, vehicular collisions, and conflict removals. The lethality of human shields may be especially intense for mesopredators, which are typically subject to harvest policies with few restrictions. Humans function as dominant "super predators" in ecosystems worldwide, resulting in human-caused mortality rates that can greatly exceed other causes of death at all trophic levels (*13*). Indeed, playback experiments indicate that the mere sound of humans induces strong avoidance by large and small carnivores alike (*14*, *15*). Mesopredators must therefore navigate competing risks from humans and large carnivores outside protected areas, which may increase the spatial heterogeneity of trophic interactions and cascading top-down effects. Whether mesopredators perceive humans as yet another competing risk to manage or as a shield protecting against large carnivores should fundamentally influence ecosystem dynamics in human-dominated systems (*16*).

Human mediation of carnivore community dynamics

Here, we examined the movements of mesopredators in relation to large-carnivore activity and human impacts to test competing hypotheses about how interactions among carnivores are moderated by human influences. If mesopredators perceive large carnivores to be a greater threat than humans, then we expected selection for human-affected areas to increase in the presence of large carnivores (human shield hypothesis). If mesopredators instead perceive humans to be a greater threat, then we expected human-affected areas to be avoided regardless of large-carnivore presence (human super predator hypothesis). We used integrated step selection functions to quantify the responses of GPS-collared coyotes (*Canis latrans*, n = 35) and bobcats (*Lynx rufus* n = 37) to wolves (*Canis lupus*, n = 22 wolves in 9 packs) and cougars (*Puma concolor*, n = 60) across rural landscapes in northern Washington, USA characterized by human impacts such as agriculture, livestock production, logging, hunting, recreation, and residential development (n = 283,536 GPS locations) (Fig. 1 and data S1 and S2) . We then used cumulative incidence function models of competing risks to estimate cause-specific mortality rates for coyotes and bobcats, examining dynamics across our two large study areas (each ~5000 km^2). Wolves began naturally

Table 1. Fates of radio-collared coyotes and bobcats. Numbers of individuals in each category are shown. The "other" category consisted of 1 accident (drowning; bobcat), 1 disease (nematodiasis; bobcat), 1 starvation or disease (bobcat), and 3 cases of intraspecific strife (2 coyotes and 1 bobcat). The human-caused category consisted of 8 bobcats killed by trappers (7 male and 1 female), 3 bobcats shot (all male), 13 coyotes shot (6 female and 7 male), and 1 coyote hit by a vehicle. Of the 11 predator-caused mortalities (including intraspecific), carcasses were consumed in 4 cases (all cougar kills; 1 cougar-killed coyote was not eaten). Two cases in which the predator species could not be identified were included in the large-carnivore category to allow the maximum mortality risk posed by large carnivores to be calculated.

Fate	Bobcats	Coyotes	Total
Large carnivore			
Black bear	0	1	1
Cougar	2	3	5
Unknown predator	1	1	2
Total large carnivore	3	5	8
Human	11	14	25
Other	4	2	6
Unknown cause of death	0	3	3
Survived/censored	19	11	30
Total	37	35	72

[1]School of Environmental and Forest Sciences, University of Washington, Seattle, WA 98195, USA. [2]Washington Department of Fish and Wildlife, Snoqualmie, WA 98065, USA. [3]Spokane Tribe of Indians, Department of Natural Resources, Wellpinit, WA 99040, USA.
*Corresponding author. Email: lprugh@uw.edu.

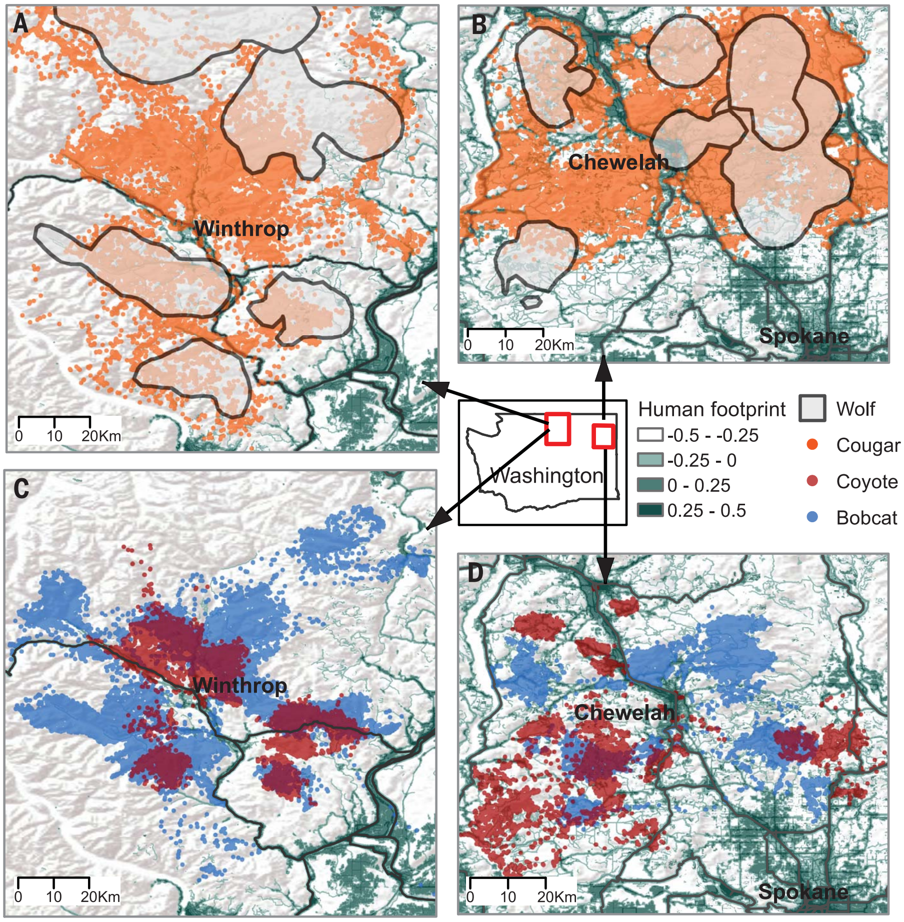


Fig. 1. Carnivore movements outside protected areas in Washington, USA. (**A** and **C**) Okanogan study area, ~5000 km^2. (**B** and **D**) Northeast study area, ~5000 km^2. The human footprint index (background) was scaled from –0.5 (corresponding to wilderness) to 0.5 (corresponding to urban areas), with a mean value of –0.33 across these landscapes. Locations of [(A) and (B)] collared large carnivores and [(C) and (D)] mesopredators are shown. All locations from 2017 to 2020 are shown for cougars (orange), bobcats (blue), and coyotes (red), and 90% kernel density polygons from summer 2020 are shown for wolf packs (gray).

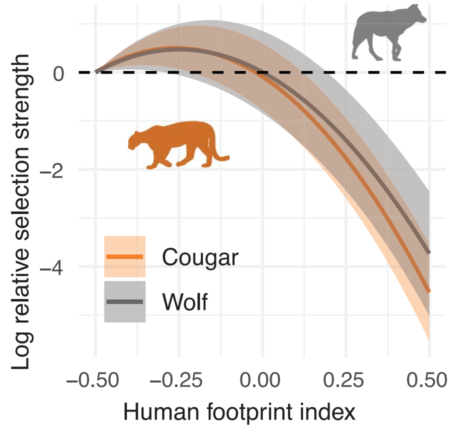


Fig. 2. Large carnivores avoid humans. Resource-selection functions show that wolves and cougars strongly avoid areas with high human footprint, thereby creating a human shield that can be used by mesopredators. Selection strength was quantified relative to areas with the lowest observed human footprint index values (human = –0.5, corresponding to wilderness), with positive values indicating selection, and negative values indicating avoidance.

recolonizing Washington in 2008 (*17*), whereas cougars were never extirpated, providing a distinct opportunity to assess the roles of recovering and established large-carnivore populations outside protected areas.

Mesopredators use a lethal human shield

Wolves and cougars strongly avoided areas with a high human footprint, thereby creating a human shield (Fig. 2 and table S1). Coyotes strongly avoided areas of high wolf use, and their selection tended to decline with increasing use by cougars (Fig. 3A and table S2). Bobcats had weaker direct responses to large carnivores, with no detectable response to wolves and a modest positive response to cougars that was likely driven by strong selection for forested habitats by both felids (Fig. 3B and table S2). However, relative use of human-affected areas by both coyotes and bobcats was higher in the presence of both wolves and cougars (Fig. 3, C to F), which supports the human shield hypothesis. Peak use by coyotes shifted from a human footprint index value of –0.03 when wolves were absent to 0.27 when wolves were present (Fig. 3C). Use of human-affected areas by coyotes was less influenced by the presence of cougars; peak use shifted from a value of -0.01 when cougars were absent to 0.10 when cougars were present (Fig. 3E). Bobcats selected far more strongly for human-affected areas when both wolves and cougars were present (table S2). In the absence of wolves and cougars, landscape use by bobcats peaked at human footprint values of -0.33 and -0.38, respectively, and then declined sharply (Fig. 3, D and F). In the presence of wolves and cougars, bobcat use remained high and did not decline until values of 0.02 and 0.08, respectively (Fig. 3, D and F). Taken together, mesopredators preferred areas with human footprint index values that were, on average, 31 percentage points higher where large carnivores were present, which corresponds to a twofold increase in human influence.

These findings indicate that mesopredators perceived large carnivores to be a greater threat than humans. Contrary to this perception, humans functioned as super predators in this system, as they were responsible for the overwhelming majority of mortality for all four species (Table 1, Fig. 4, and supplementary text). The annual risk of mortality from humans was three times higher than the risk from large carnivores for coyotes ($-x$ = 0.30 versus 0.10) (Fig. 4) and 3.8 times higher for bobcats ($-x$ = 0.25 versus 0.07) (Fig. 4).

Synergistic effects of large carnivores and humans

Despite global increases in the extent of protected area networks, anthropogenic impacts on natural systems are pervasive and continuing to expand (*3*, *18*). Wildlife that persist in human-altered areas must therefore navigate a complex suite of threats and resources, which can be particularly challenging for carnivores because humans have an especially low tolerance for this guild (*19*). Here, we show that mesopredators use a human shield response to risks posed by large carnivores outside protected areas, increasing their selection for human-altered areas where large carnivores are present. Unlike in protected areas, however, these shields are lethal in most of the land area where carnivore populations persist (*3*, *4*). In our system, humans

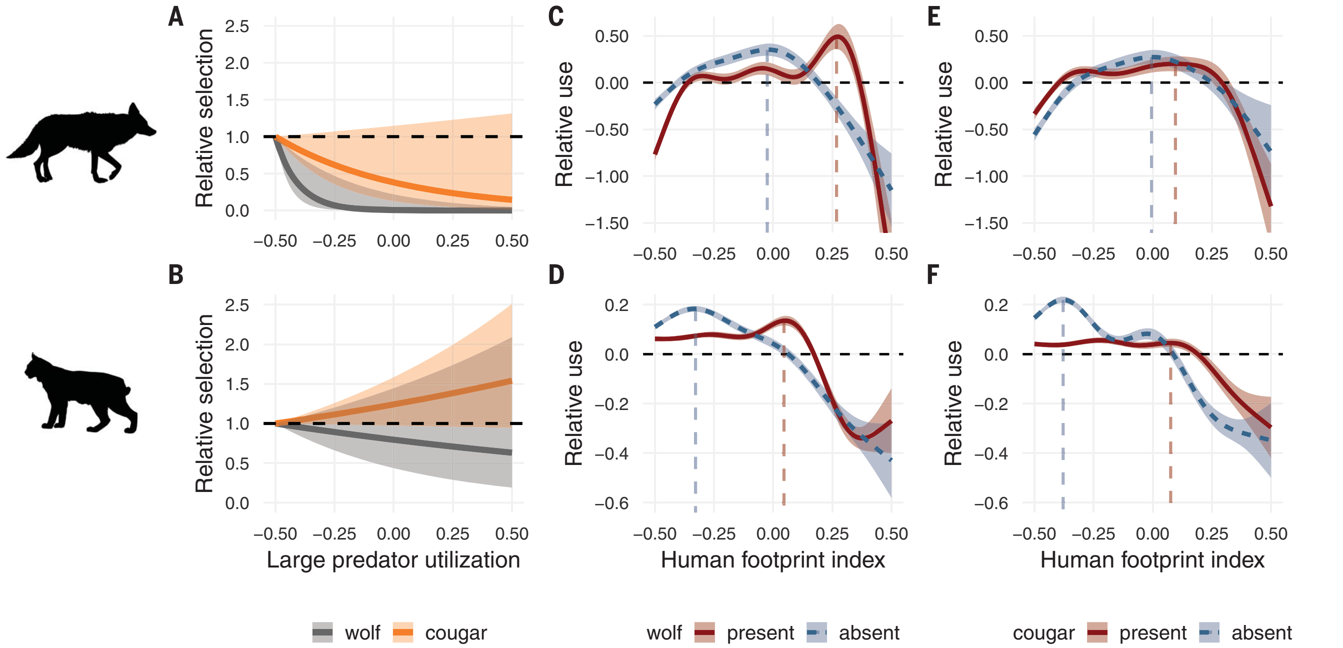
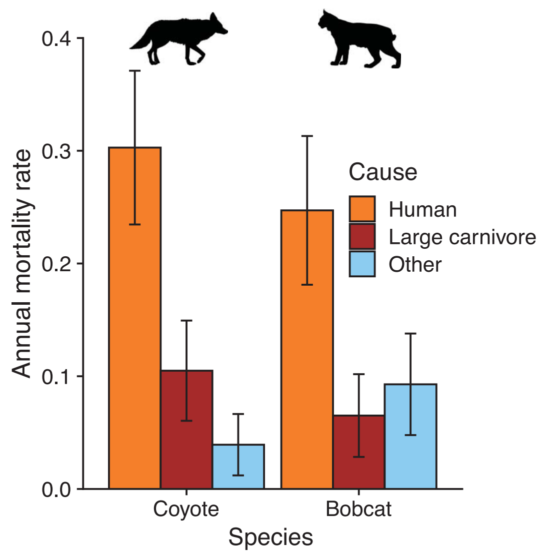


Fig. 3. Responses of mesopredators to humans and large carnivores. Relative selection strength [± 95% confidence interval (CI)] by (**A**) coyotes and (**B**) bobcats in response to large carnivores shows stronger avoidance by coyotes. Selection was quantified relative to the minimum observed value of large-carnivore utilization, corresponding to the absence of large carnivores. Values above 1 indicate selection and values below 1 indicate avoidance. (**C** to **F**) Average-effect plots show the responses of [(C) and (E)] coyotes and [(D) and (F)] bobcats to the human footprint index in the presence (red solid lines) and absence (blue dashed lines) of [(C) and (D)] wolves and [(E) and (F)] cougars (± 95% CI). Average-effect plots account for covariation among predictors by quantifying the relative change in expected use across the range of the focal variable while averaging over the observed values of other variables. Positive values indicate above-average use (on the logarithmic scale) and negative values indicate below-average use. Vertical dashed lines indicate peak use with respect to the human footprint index in the presence and absence of large carnivores. Wolf and cougar presence was determined based on home-range boundaries delineated by utilization distribution models.

Fig. 4. Patterns of mesopredator mortality. Rates of mortality caused by humans (orange), large carnivores (brown), and other sources (blue) estimated from radio-collared bobcats (n = 37 individuals, 59 animal years) and coyotes (n = 32 individuals, 52 animal years). Mean annual mortality rates and standard error bars are shown. See Table 1 and data S1 for detailed information about causes of mortality.



were approximately three times more lethal for mesopredators than were large carnivores, indicating that fear of large carnivores may drive mesopredators into areas of even greater risk from humans. These results indicate that large carnivores can directly and indirectly perform ecosystem functions such as mesopredator suppression outside protected areas, and their impacts in human-dominated systems may paradoxically be amplified because of reduced enemy-free space and synergies with human super predators.

Multiple predator species in a system can lead to emergent effects on prey, whereby mutual interference can reduce hunting efficiency, or conversely, lead to superadditive predation rates by reducing available refuges (*6*, *20*, *21*). Our study was conducted across two large regions characterized by a diverse large carnivore guild despite strong human influences. Mesopredators faced risks from wolves, cougars, and black bears (*Ursus americanus*) in addition to humans. Although bobcats and coyotes responded similarly to both wolves and cougars by increasing their use of human-affected areas, bobcats avoided human-affected areas far more strongly in the absence of large carnivores, thus leaving a greater capacity for increased use of these areas in the presence of large carnivores. The synergistic effects of humans and large carnivores may therefore vary substantially in strength across species, highlighting the importance of multispecies studies in understanding carnivore community dynamics (*22*).

Although use of a lethal human shield may seem maladaptive, there are two likely explanations for this behavior that are not mutually exclusive. First, wildlife may be less equipped to accurately evaluate risk from contemporary humans compared with other coevolved predators (*23*), leading to behaviors that increase their risk of anthropogenic mortality. For example, modern firearms facilitate hunting by humans from distances where visual, olfactory, and auditory cues of risk may be absent. Likewise, the use of equipment such as traps and

snares minimizes direct cues of risk because humans are not present at the time of capture. Wildlife may thus respond more strongly to direct cues of human presence, such as the sounds of humans talking that are used in playback experiments (*14*, *15*, *24*), than to the diffuse and heterogeneous cues that are present in human-affected areas. Second, subordinate species may benefit from increased access to anthropogenic resources such as refuse, livestock, and crops (*25*, *26*), and these benefits could offset some or all of the fitness costs associated with increased mortality risk. Whereas our analyses focused on understanding the risk landscape, the distribution of resources such as anthropogenic subsidies and scavenging opportunities should strongly affect carnivore community dynamics outside protected areas as well.

Indirect risk effects from antithetical threats

The fitness costs of antipredator behavior (i.e., risk effects) are often measured in terms of reduced foraging opportunities, physiological stress, or altered movement patterns (*27*). Our findings show that antipredator behavior can also increase risk of mortality from an opposing threat. Mortality risks posed by large carnivores and humans are not simply competing (i.e., an individual can die from one or the other, but not both); these risks are antithetical, because antipredator behavior to avoid one threat puts the individual at increased risk from the other. These indirect risk effects may be especially important in human-dominated areas where risk cues from humans can be difficult for wildlife to accurately gauge (*23*). Our findings highlight the importance of spatial dynamics in landscapes of fear, but temporal shifts in activity patterns of subordinate species could also lead to indirect risk effects (*28*). For example, Shores *et al*. (*29*) found that recolonizing wolves caused a diurnal shift in coyote activity that increased their temporal overlap with humans, which may further increase their risk. Conversely, many wildlife species shift toward increased nocturnality in response to human activity, which could increase risk from large carnivores (*30*). The spatial and temporal responses to antithetical threats should thus combine to determine the net strength of indirect risk effects (*31*).

The patterns of mortality we documented were inversely proportional to the strength of avoidance behaviors. For example, coyotes had the strongest avoidance response to wolves, followed by cougars and lastly, humans (table S2). Yet, no collared coyotes were known to have been killed by wolves in our study, whereas 3 were killed by cougars and 14 were killed by humans (13 were shot and 1 was hit by a vehicle) (Table 1). This discrepancy highlights the importance of considering the nonlethal effects of predators along with the lethal effects, because patterns of behavior or mortality alone may not accurately reflect the relative importance of competing threats in limiting population growth (*24*).

Carnivore conservation in anthropogenic landscapes

The release of mesopredator populations in the wake of large-carnivore extirpations has led to negative societal and ecological outcomes throughout the world (*1*, *32*). Amplified suppression by means of large carnivores and lethal human shields may help to limit problems caused by mesopredator overabundance, but whether this intense top-down control is ultimately beneficial remains a question. A far greater proportion of small-carnivore species are ranked as "least concern" by the International Union for Conservation of Nature (IUCN) compared with large carnivore species (64 versus 22%), but populations of both large and small carnivores are declining at alarming rates worldwide (*33*, *34*). Thus, recovery of large carnivores in anthropogenic landscapes could lead to unsustainably high levels of persecution and threaten some mesopredator populations with local extirpation. Although mesopredators such as coyotes and jackals (*Canis* spp.) have proven to be notably resilient when faced with intense persecution, this resilience is driven by a high capacity for compensatory reproduction and immigration that might not be possible for some other species (*35*–*37*). Urban-wildland gradients, where the paradox of the lethal human shield should be most influential, should thus be priority regions for monitoring trends of species that are subject to mortality from both humans and large carnivores.

Large-carnivore populations are recovering in areas throughout North America and Europe (*38*), yet the landscapes to which they are returning are vastly different than those in which they once thrived. Carnivores are often intensively managed and held to levels far below their carrying capacities outside protected areas, which calls into question whether they can perform the ecosystem functions of prey and mesopredator regulation that have been documented within protected areas (*16*). This question is critical to answer, because the importance of large carnivores for healthy ecosystems is often used as a justification to promote their recovery despite the societal costs incurred (*39*). Wolves began returning to Washington less than two decades ago, and our findings indicate that they are already affecting coyote and bobcat populations, but through behavioral effects more than direct mortality. Cougars were never extirpated from Washington, and our findings indicate that they affect coyotes and bobcats through both behavioral effects and direct mortality, despite being intensively managed themselves. As landscapes of coexistence between large carnivores and humans expand with recovery and restoration efforts (*40*, *41*), the paradox of the lethal human shield may become an increasingly dominant driver of carnivore community dynamics.

REFERENCES AND NOTES

1. W. J. Ripple *et al*., *Science* **343**, 1241484 (2014).
2. C. Wolf, W. J. Ripple, *R. Soc. Open Sci.* **4**, 170052–170052 (2017).
3. M. Pacifici, M. Di Marco, J. E. M. Watson, *Conserv. Lett.* **13**, e12748 (2020).
4. R. Woodroffe, J. R. Ginsberg, *Science* **280**, 2126–2128 (1998).
5. S. M. Durant *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **114**, 528–533 (2017).
6. L. R. Prugh, K. J. Sivy, *Ecol. Lett.* **23**, 902–918 (2020).
7. W. M. Sarmento, J. Berger, *Biol. Conserv.* **212**, 316–326 (2017).
8. J. Berger, *Biol. Lett.* **3**, 620–623 (2007).
9. R. J. Moll *et al*., *Urban Ecosyst.* **21**, 765–778 (2018).
10. T. B. Muhly, C. Semeniuk, A. Massolo, L. Hickman, M. Musiani, *PLOS ONE* **6**, e17050 (2011).
11. S. Steyaert *et al*., *Proc. R. Soc. Biol. Sci. Ser. B* **283**, 20160906 (2016).
12. J. F. Goldberg, M. Hebblewhite, J. Bardsley, *PLOS ONE* **9**, e91417 (2014).
13. C. T. Darimont, C. H. Fox, H. M. Bryan, T. E. Reimchen, *Science* **349**, 858–860 (2015).
14. J. P. Suraci, M. Clinchy, L. Y. Zanette, C. C. Wilmers, *Ecol. Lett.* **22**, 1578–1586 (2019).
15. J. A. Smith *et al*., *Proc. R. Soc. Biol. Sci. Ser. B* **284**, 20170433 (2017).
16. D. P. J. Kuijper *et al*, *Proc. R. Soc. Biol. Sci. Ser. B* **283**, 20161625 (2016).
17. Washington Department of Fish and Wildlife, Confederated Tribes of the Colville Reservation, Spokane Tribe of Indians, USDA-APHIS Wildlife Services, U.S. Fish and Wildlife Service, "Washington Gray Wolf Conservation and Management 2021 Annual Report," (Washington Department of Fish and Wildlife, Ellensburg, WA, USA, 2022).
18. R. Dirzo *et al*., *Science* **345**, 401–406 (2014).
19. J. T. Bruskotter, R. S. Wilson, *Conserv. Lett.* **7**, 158–165 (2014).
20. A. Sih, G. Englund, D. Wooster, *Trends Ecol. Evol.* **13**, 350–355 (1998).
21. M. J. Jeffries, J. H. Lawton, *Biol. J. Linn. Soc. Lond.* **23**, 269–286 (1984).
22. R. A. Montgomery, R. J. Moll, E. Say-Sallaz, M. Valeix, L. R. Prugh, *Biol. Conserv.* **233**, 1–11 (2019).
23. A. C. Nisi, J. F. Benson, C. C. Wilmers, *Oikos* **2022**, e09051 (2022).
24. M. C. Allen, M. Clinchy, L. Y. Zanette, *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2112404119 (2022).
25. T. M. Newsome *et al*., *Glob. Ecol. Biogeogr.* **24**, 1–11 (2015).
26. M. A. Parsons, T. M. Newsome, J. K. Young, *Front. Ecol. Environ.* **20**, 31–39 (2022).
27. L. R. Prugh *et al*., *Biol. Conserv.* **232**, 194–207 (2019).
28. M. S. Palmer *et al*., *Trends Ecol. Evol.* **37**, 911–925 (2022).
29. C. R. Shores, J. A. Dellinger, E. S. Newkirk, S. M. Kachel, A. J. Wirsing, *Behav. Ecol.* **30**, 1324–1335 (2019).
30. K. M. Gaynor, C. E. Hojnowski, N. H. Carter, J. S. Brashares, *Science* **360**, 1232–1235 (2018).
31. A. Van Scoyoc, J. A. Smith, K. M. Gaynor, K. Barker, J. S. Brashares, *J. Anim. Ecol.* 1365-2656.13892 (2023).
32. L. R. Prugh *et al*., *Bioscience* **59**, 779–791 (2009).
33. C. Marneweck *et al*., *Biol. Conserv.* **255**, 109005 (2021).
34. J. Fernández-Sepúlveda, C. A. Martín, *Mamm. Biol.* **102**, 1911–1925 (2022).
35. G. Péron, *J. Anim. Ecol.* **82**, 408–417 (2013).
36. L. Minnie, A. Gaylard, G. I. H. Kerley, *J. Appl. Ecol.* **53**, 379–387 (2016).
37. J. C. Kilgo, C. E. Shaw, M. Vukovich, M. J. Conroy, C. Ruth, *J. Wildl. Manage.* **81**, 1386–1393 (2017).
38. G. Chapron *et al*., *Science* **346**, 1517–1519 (2014).
39. E. G. Ritchie *et al*., *Trends Ecol. Evol.* **27**, 265–271 (2012).
40. A. Oriol-Cotterill, M. Valeix, L. G. Frank, C. Riginos, D. W. Macdonald, *Oikos* **124**, 1263–1273 (2015).
41. B. Gehr *et al*., *Oikos* **126**, 1389–1399 (2017).

42. L. R. Prugh, GPS tracking of bobcats and coyotes in northern Washington. Movebank Data Repository (2023); https://doi.org/10.5441/5001/5441.gm93267b.

ACKNOWLEDGMENTS

This research was made possible through the efforts of numerous dedicated field personnel, especially C. Vanbianchi and A. Machowicz. Wolf captures were led by T. Roussin (Washington Department of Fish and Wildlife), B. Maletzke (Washington Department of Fish and Wildlife), and D. Abrahamson (Spokane Tribe). L. Satterfield led cougar captures in the Okanogan. J. Rohrer assisted with field housing and logistics. C. Darimont and an anonymous reviewer provided valuable feedback on earlier versions of this manuscript. **Funding:** Funding was provided by the National Science Foundation grant DEB-1652420 (L.R.P.); Washington Department of Fish and Wildlife grant 16-07846 (A.J.W. and L.R.P.); and Fulbright Australia Program (C.X.C.). **Author contributions:** Conceptualization: L.R.P.; Methodology: L.R.P., C.X.C., A.J.W., and B.N.K.; Investigation: R.M.W., L.R.P., T.R.G., S.L.W., and B.N.K.; Visualization: C.X.C. and L.R.P.; Writing – original draft: L.R.P. and C.X.C.; Writing – review and editing: L.R.P., C.X.C., R.M.W., T.R.G., A.J.W., B.N.K., and S.L.W. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Data and code needed to reproduce the findings are available in the supplementary materials. Bobcat and coyote location data are publicly archived in the Movebank Data Repository (*42*): https://doi.org/10.5441/5001/5441.gm93267b. Data for wolves and cougars is not freely available because of conservation concerns. Data will be made available, upon request to the authors, to qualified researchers who agree in writing not to publish or share exact locations or other information that could threaten or lead to harm of the animals. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adf2472
Materials and Methods
Supplementary Text
Fig. S1
Tables S1 to S4
References (*43*–*50*)
MDAR Reproducibility Checklist
Data S1 and S2
Code S1 to S3

Submitted 8 October 2022; accepted 22 March 2023
10.1126/science.adf2472

HEART REGENERATION

Interplay between calcium and sarcomeres directs cardiomyocyte maturation during regeneration

Phong D. Nguyen[1]*, Iris Gooijers[1]†, Giulia Campostrini[2]†, Arie O. Verkerk[3,4], Hessel Honkoop[1], Mara Bouwman[1], Dennis E. M. de Bakker[1,5], Tim Koopmans[1,6], Aryan Vink[7], Gerda E. M. Lamers[8], Avraham Shakked[9], Jonas Mars[1], Aat A. Mulder[10], Sonja Chocron[1], Kerstin Bartscherer[1,6], Eldad Tzahor[9], Christine L. Mummery[2], Teun P. de Boer[11], Milena Bellin[2,12,13], Jeroen Bakkers[1,14]*

Zebrafish hearts can regenerate by replacing damaged tissue with new cardiomyocytes. Although the steps leading up to the proliferation of surviving cardiomyocytes have been extensively studied, little is known about the mechanisms that control proliferation and redifferentiation to a mature state. We found that the cardiac dyad, a structure that regulates calcium handling and excitation-contraction coupling, played a key role in the redifferentiation process. A component of the cardiac dyad called leucine-rich repeat–containing 10 (Lrrc10) acted as a negative regulator of proliferation, prevented cardiomegaly, and induced redifferentiation. We found that its function was conserved in mammalian cardiomyocytes. This study highlights the importance of the underlying mechanisms required for heart regeneration and their application to the generation of fully functional cardiomyocytes.

After a myocardial infarction (MI), the mammalian heart loses millions of cardiomyocytes (CMs) that are replaced by a permanent fibrotic scar. Inducing endogenous CM proliferation to replace the lost CMs is a promising strategy (*1*) that is observed after injury in zebrafish (*2*), salamanders (*3*), and embryonic or neonatal mammalian hearts (*4*–*6*). However, an underappreciated aspect of heart regeneration is the control of CM proliferation. Successful induction of CM proliferation (*7*–*11*) results in cardiomegaly (large heart) owing to uncontrolled CM proliferation despite positive regenerative outcomes. Furthermore, CM proliferation is preceded by CM dedifferentiation, which results in sarcomere breakdown, metabolic reprogramming, and reexpression of embryonic genes (*2*, *12*–*18*). If this dedifferentiated and proliferative state continues, long-term negative outcomes can reduce CM integration into the heart and decrease heart function (*19*–*23*).

The zebrafish has a natural regenerative capacity, which makes it an excellent model to investigate how CM proliferation and dedifferentiation-redifferentiation is controlled (*2*, *12*–*15*, *24*–*26*). However, little is known about the mechanisms that restore the injured heart to its original size and function.

Ca^{2+} handling of border-zone CMs changes during regeneration

CMs exhibit cycles of excitation-contraction (EC) coupling and Ca^{2+} handling (*27*, *28*). During EC coupling, an influx of Ca^{2+} through L-type calcium channels (LTCCs) triggers Ca^{2+} release from the sarcoplasmic reticulum (SR). This results in sarcomere contraction, after which Ca^{2+} is shuttled back into the SR or out of the cell. Because Ca^{2+} handling changes as CMs mature, we developed an ex vivo imaging system using a transgenic zebrafish line that expressed the fluorescent Ca^{2+} sensor GCaMP6f (*29*) specifically in CMs to track the dynamics during regeneration (*26*) (Fig. 1, A to A").

Comparisons between CMs from the uninjured remote zone (RZ) and the border zone (BZ)—the region adjacent to the infarcted area, where CMs dedifferentiate and proliferate—showed differences in Ca^{2+} influx and efflux duration during the early regeneration time points, which were resolved by 21 days post-injury (dpi) (Fig. 1, B and C). There were no differences between trabecular and cortical BZ-CMs despite their reported different regenerative potential (*25*, *30*) (fig. S1, A and A'). Next, we compared BZ-CM Ca^{2+} transients with embryonic CMs and found that although Ca^{2+} influx was indistinguishable between time points (fig. S1B), there were significant differences in the Ca^{2+} efflux between embryonic and initial stages of regeneration compared with later regeneration stages (Fig. 1D). Thus, BZ-CMs functionally reverted to an embryonic-like state during the early regeneration stages in a time-dependent manner, which agrees with previous histological and transcriptomic observations (*12*, *13*).

Lrrc10 marks the onset of CM maturation

We next used single-cell RNA sequencing (scRNA-seq) on BZ-CMs from early (7 dpi) and late (21 dpi) regeneration stages to more broadly characterize the process of redifferentiation and maturation (table S1). Initial analysis revealed that CMs clustered in a ring-like manner (Fig. 2A and fig. S2, A and A'), which mimicked a dedifferentiation-redifferentiation cycle when analyzed using CytoTrace (*31*) (Fig. 2B). To simplify downstream analysis, cells were projected into a linear pseudotime trajectory (*32*) to represent the regenerating CM axis (Fig. 2C and fig. S2, B to B"). In this axis, we determined the order of processes that CMs undergo during regeneration and broadly categorized this axis into an early and a late regeneration phase (Fig. 2D). The early regeneration phase showed high expression of genes related to injury, heart development, sarcomeres, cell cycle, glycolysis, and mitochondrial-related translation (Fig. 2D and fig. S2C), consistent with the previously identified CM dedifferentiation phase (*2*, *12*–*18*, *24*, *25*, *33*, *34*). In the late regeneration phase, CMs progressed through states associated with CM maturity, such as Ca^{2+} regulation, mitochondrial oxidative phosphorylation (OXPHOS), cytoplasmic-related translation, and ion channels (Fig. 2D and fig. S2C').

Given that Ca^{2+} dynamics were different between early and late regeneration phases and that genes related to Ca^{2+} regulation were prominently expressed during the late regeneration phase, we searched for candidate genes that could potentially be involved in this state of the regenerating CM axis. An interpopulation analysis of Mexican cavefish, in which heart regeneration does not take place after injury, identified the leucine-rich repeat–containing 10 (*lrrc10*) gene—encoding an LTCC accessory protein (*35*, *36*)—as a player in heart regeneration (*37*); yet, the underlying mechanism remains unknown. Specifically, *lrrc10* knockout zebrafish mutants retain a scar (*37*) (fig. S3, A and A') yet display normal BZ-CM proliferation (*37*). Notably, *lrrc10* was down-regulated in the early regeneration phase and up-regulated at the start of the late regeneration phase (Fig. 2D and fig. S2, C and C'), which was experimentally validated (fig. S4, A to G). Furthermore, comparisons between uninjured adult *lrrc10* siblings revealed no

[1]Hubrecht Institute-KNAW and University Medical Center Utrecht, Utrecht, Netherlands. [2]Department of Anatomy and Embryology, Leiden University Medical Center, Leiden, Netherlands. [3]Department of Medical Biology, Amsterdam Cardiovascular Sciences, University of Amsterdam, Amsterdam University Medical Center, Amsterdam, Netherlands. [4]Department of Experimental Cardiology, University of Amsterdam, Amsterdam University Medical Center, Amsterdam, Netherlands. [5]Leibniz Institute on Aging – Fritz Lipmann Institute (FLI), Jena, Germany. [6]Department of Animal Physiology, Osnabrueck University, Osnabrück, Germany. [7]Department of Pathology, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands. [8]Core Facility Microscopy, Institute of Biology, Leiden University, Leiden, Netherlands. [9]Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot, Israel. [10]Department of Cell and Chemical Biology, Leiden University Medical Center, Leiden, Netherlands. [11]Department of Medical Physiology, Division of Heart and Lungs, University Medical Center Utrecht, Utrecht, Netherlands. [12]Department of Biology, University of Padua, Padua, Italy. [13]Veneto Institute of Molecular Medicine, Padua, Italy. [14]Department of Pediatric Cardiology, Division of Pediatrics, University Medical Center Utrecht, Utrecht, Netherlands.
***Corresponding author. Email: p.nguyen@hubrecht.eu (P.D.N.); j.bakkers@hubrecht.eu (J.B.)**
†These authors contributed equally to this work.

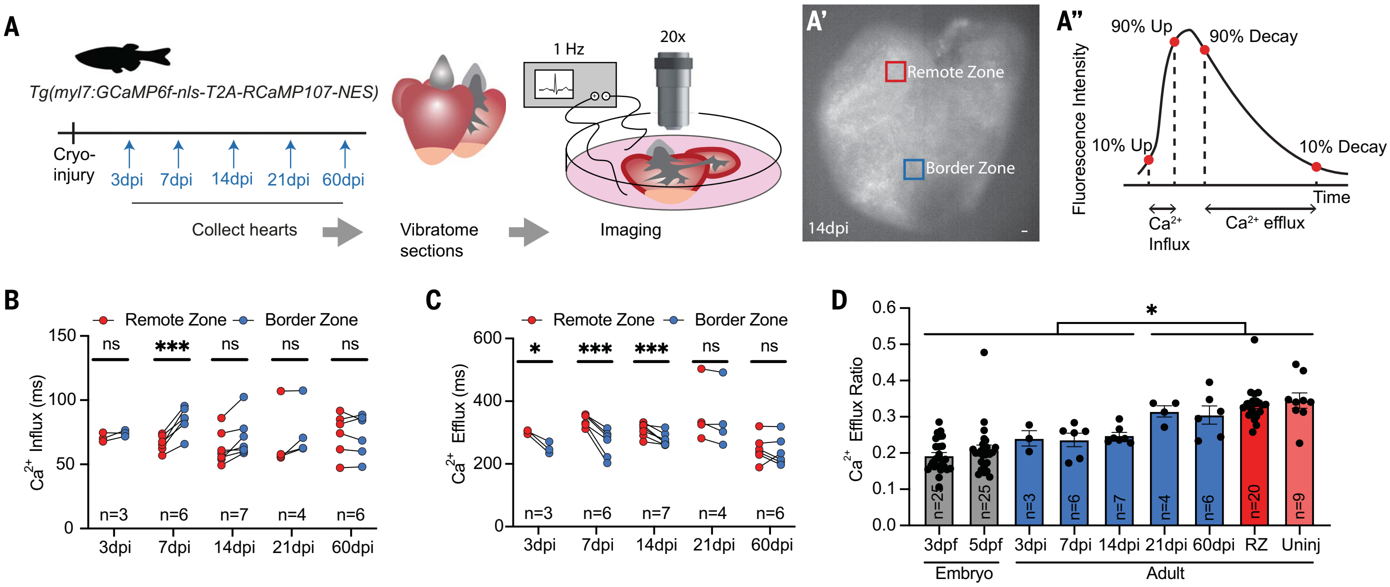


Fig. 1. Ca^{2+} handling matures in CMs during regeneration. (**A'** and **A"**) Strategy for monitoring live Ca^{2+} cycling in adult hearts. (**A**) Adult fish containing the Ca^{2+} sensors *Tg*(*myl7:GCaMP6f-nls-T2A-RCaMP107-NES-pA*) were injured. At a designated time point, cardiac slices were made and attached to electrodes for synchronous electrical stimulation, and the GCaMP signal was recorded at high speed ex vivo. (A') The fluorescence intensity changes over time in the RZ (red) and BZ (blue) were extracted from recordings. Scale bar, 100 μm. (A") Example of the GCaMP peak whereby various characteristics of the curve were analyzed with custom MATLAB scripts. The difference in duration between 10% and 90% upstroke corresponded to Ca^{2+} influx. The difference in duration between 90% and 10% decay corresponded to Ca^{2+} efflux. (**B**) Quantification of Ca^{2+} influx of adult hearts after injury reveals only a significant difference in Ca^{2+} handling between RZ- and BZ-CMs at 7 dpi. Paired *t* test was conducted. (**C**) Quantification of Ca^{2+} efflux of adult hearts after injury reveals that Ca^{2+} handling is different between RZ- and BZ-CMs during early stages of regeneration (3 dpi, 7 dpi, and 14 dpi) with a recovery from 21 dpi onward. Paired *t* test was conducted. (**D**) Quantification of Ca^{2+} efflux of both embryonic [3 days postfertilization (dpf) and 5 dpf] and adult (RZ and uninjured) hearts reveal that BZ-CMs at early regeneration stages (3 dpi, 7 dpi, and 14 dpi) functionally resemble embryonic CMs, whereas they resemble adult CMs at late (21 dpi and 60 dpi) regeneration stages. One-way analysis of variance (ANOVA) was conducted. Throughout the figures, *P* values are as follows: *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$; ns, not significant.

differences in Ca^{2+} influx; however, Ca^{2+} efflux was prolonged (fig. S3B). During regeneration, *lrrc10*−/− CMs failed to recover both their Ca^{2+} influx and efflux at the early (7 dpi) and even at the late regeneration stages (21 dpi and 60 dpi) (Fig. 2, E to F, and fig. S3C). Moreover, *lrrc10*−/− BZ-CMs functionally maintained an embryonic Ca^{2+} handling phenotype even at the late regeneration stage (60 dpi), indicating that they failed to redifferentiate (Fig. 2G and fig. S3D). These results suggest a role for Lrrc10 during CM maturation. Corroborating this model, the expression of *lrrc10* in BZ-CMs negatively correlated with the expression of dedifferentiation markers (embryonic MYH and pS6) (Fig. 2, H and H', and fig. S5, A and A'). Additionally, expression of dedifferentiation markers (*myomesin*, *hexokinase1*, and pS6) reverted to normal levels in *lrrc10*+/+ hearts at the late regeneration time point (21 dpi), whereas these markers remained elevated in *lrrc10*−/− BZ-CMs (Fig. 2, I and I', and fig. S5, B to C"). By contrast, Lrrc10 overexpression (OE) at the early regeneration time point (7 dpi) eliminated pS6 expression specifically in BZ-CMs (fig. S5, D and D'). Thus, Lrrc10 plays a critical role in actively maturing CMs by inhibiting early regeneration hallmarks, whereas *lrrc10* mutants appeared to be perpetually held in an early regeneration phase and consequently failed to regenerate.

LRRC10 OE can induce maturation in human-induced pluripotent stem cell (hiPSC)–CMs

Human patients and mice harboring homozygous loss-of-function mutations in *LRRC10* and *Lrrc10*, respectively, develop dilated cardiomyopathy (*35*, *38*, *39*). To test whether LRRC10 could also induce CM maturation in humans, we used hiPSC-CMs because they resemble fetal CMs (*28*, *40*). hiPSC-CMs with LRRC10 OE displayed significantly improved sarcomere organization and sarcomere length (Fig. 3, A to A"). Furthermore, we observed a significant increase in LTCC ($Ca_v1.2$) expression after LRRC10 OE (Fig. 3, B and B') as well as increased connexin-43 (CX43) expression and localization between neighboring CMs (fig. S6, A to B'). Given the increase in an LTCC component as well as intracellular coupling, Ca^{2+} transients of hiPSC-CM cultures were measured. Both Ca^{2+} influx and efflux duration showed a significant increase in cultures with LRRC10 OE (fig. S6C), mirroring the dynamics seen in three-dimensional (3D) hiPSC cardiac microtissues, which display a mature CM state (*40*) (fig. S6D). Moreover, LRRC10 OE CM cultures had significantly more transverse tubules (t-tubules) compared with the immature hiPSC-CMs (fig. S6, E and E').

To assess whether LRRC10 OE induced a broader maturation signature, we analyzed transcription profiles by scRNA-seq. We identified six distinct clusters, and application of CytoTrace (*31*) and pseudotime trajectory analysis to determine the maturation axis showed that hiPSC-CM maturation correlated with LRRC10 expression levels (Fig. 3, C to E; fig. S7, A to B"; and table S1). We performed further examination of the maturation status using markers such as sarcomeric isoform switching (Fig. 3F and fig. S7C), metabolic state switching (Fig. 3G), expression of EC coupling markers (fig. S7D), and comparison with expression profiles of hiPSC-CMs matured in a 3D microtissue format (fig. S7E). Overall, we found that the expression of LRRC10 is not associated with the end of maturation, but rather the onset of its expression appears to advance progression of hiPSC-CMs to a more mature state.

The cardiac dyad is disrupted after injury and requires Lrrc10 to reassemble

The cardiac dyad is a region that assists in efficient EC coupling by placing components that regulate Ca^{2+} flux (LTCC complex and SR) and sarcomeres in proximity. Because Lrrc10

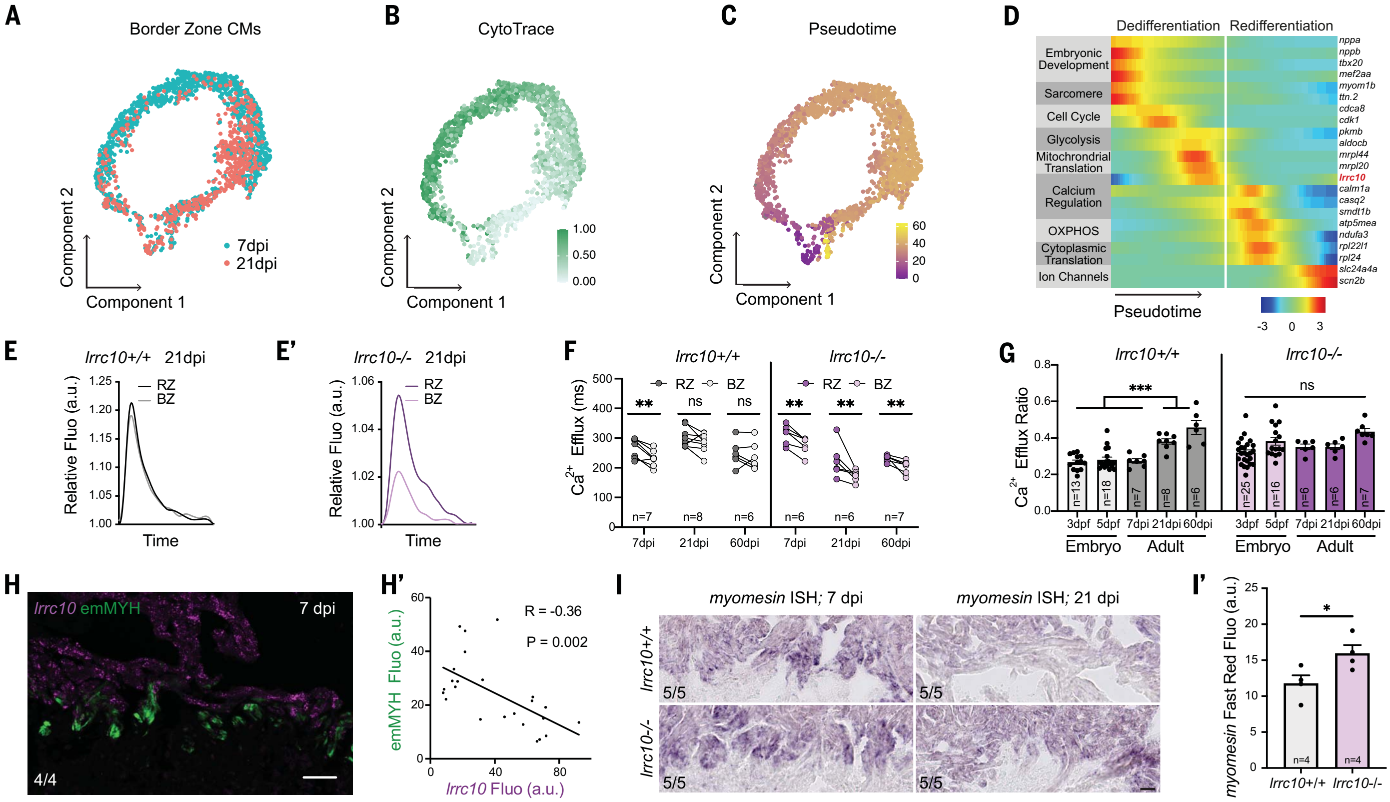


Fig. 2. Ca^{2+} handling maturation requires *lrrc10*. (**A**) Uniform manifold approximation and projection (UMAP) representation of the scRNA-seq data obtained from 7 dpi and 21 dpi BZ-CMs. (**B**) UMAP representation of scaled CytoTrace scores, which indicates the extent of differentiation. A value of 0 (white) is the most differentiated, and a value of 1 (green) is the least differentiated. (**C**) UMAP representation of linear pseudotime trajectory. A value of 0 (purple) indicates the start of pseudotime, whereas a value of 60 (yellow) indicates the end of pseudotime. (**D**) Heatmaps of gene expression that are grouped into processes across pseudotime. The red gene is *lrrc10*, and the white horizontal line represents the split between the de- and redifferentiation phases. (**E**) Relative GCaMP fluorescence intensity profile of *lrrc10+/+* in RZ- and BZ-CMs at 21 dpi. a.u., arbitrary units. (**E'**) Relative GCaMP fluorescence intensity profile of *lrrc10−/−* in RZ- and BZ-CMs at 21 dpi. (**F**) Quantification of Ca^{2+} efflux in *lrrc10+/+* and *lrrc10−/−* at an early regeneration time point, 7 dpi, and late regeneration time points of 21 dpi and 60 dpi. Paired *t* test was conducted. (**G**) Quantification of the Ca^{2+} efflux in embryonic and adult CMs in *lrrc10* siblings reveals that *lrrc10−/−* functionally resemble embryonic CMs. One-way ANOVA was conducted. (**H**) RNAscope with *lrrc10* probe (purple) and embryonic myosin (emMYH) antibody (green) at the BZ at 7 dpi. Number indicates biological replicates, and the scale bar is 50 μm. (**H'**) Quantification of *lrrc10* RNAscope (*x* axis, purple) and emMYH (*y* axis, green) signals reveal a negative correlation of the fluorescence intensity. R is the R^2 value, and P is the *P* value. (**I**) In situ hybridization with *myomesin* probe marking BZ-CM expression at 7 dpi in *lrrc10+/+* and *lrrc10−/−*. Number represents biological replicates, and the scale bar is 100 μm. (**I'**) Quantification of FastRed *myomesin* signal at 21 dpi reveals significantly higher expression in *lrrc10−/−* BZ-CMs. Number represents biological replicates, and a *t* test was conducted.

associates with the LTCC complex (*35*, *36*), we conducted patch-clamp measurements on isolated *lrrc10* wild-type and mutant adult zebrafish CMs to examine whether the aberrant Ca^{2+} transient behavior could be explained by altered LTCC-generated current ($I_{Ca,L}$) properties. These experiments indicated a decrease in $I_{Ca,L}$ density (Fig. 4A), which may have resulted in a smaller influx of Ca^{2+} ions through $I_{Ca,L}$ channels. However, this was counteracted by incomplete activation and Ca^{2+}-dependent facilitation (*41*) (Fig. 4A') and prolonged $I_{Ca,L}$ (Fig. 4A''), ultimately resulting in action potential prolongation (fig. S8, A to A'''). These effects presumably reflected the downstream *lrrc10*-dependent Ca^{2+} handling dynamics.

Mammalian LRRC10 colocalizes with dyad components, such as α-actinin (*39*, *42*, *43*), RyR2 (*43*), and junctophilin 2 (*36*), and binds to sarcomeres via α-actinin (*39*, *42*, *43*) and the LTCC complex via $Ca_v1.2$ (*35*). Because it is unclear whether a zebrafish cardiac dyad exists, we investigated the localization of its components in uninjured adult CMs. All the dyad components, in addition to green fluorescent protein (GFP)–Lrrc10 were coexpressed and localized to the z-disc (Fig. 4B and fig. S8, B to B''), thereby identifying the zebrafish equivalent cardiac dyad. We next examined the distributions of these dyad components in *lrrc10* siblings. As expected, uninjured *lrrc10+/+* CMs showed clear colocalization of the dyad components (fig. S8C). However, although uninjured *lrrc10−/−* CMs displayed comparable localization of sarcomeres (α-actinin), the LTCC complex ($Ca_v1.2$) appeared disorganized (fig. S8C'), further complementing the Ca^{2+} flux and patch-clamp data. In the context of regeneration, both *lrrc10* wild-type and mutant siblings displayed disorganization of both α-actinin and $Ca_v1.2$ at the early regeneration time point, 7 dpi (Fig. 4C). At later regeneration time points (21 dpi and 60 dpi), and supporting the Ca^{2+} flux analysis, *lrrc10+/+* BZ-CMs had reorganized their dyad components such that they were comparable to those of uninjured hearts. On the other hand, although *lrrc10−/−* BZ-CMs at 21 dpi and 60 dpi showed sarcomere reorganization, they failed to relocalize $Ca_v1.2$ to the z-disc (Fig. 4D and fig. S8, D and D'). We next investigated whether the disorganization of the cardiac dyad after injury was also observed in mammals. In MI-injured mouse hearts, BZ-CMs adjacent to the scar showed a complete disorganization of the dyad at both early and late time points (Fig. 4, E and F, and fig. S9, A to B'), and

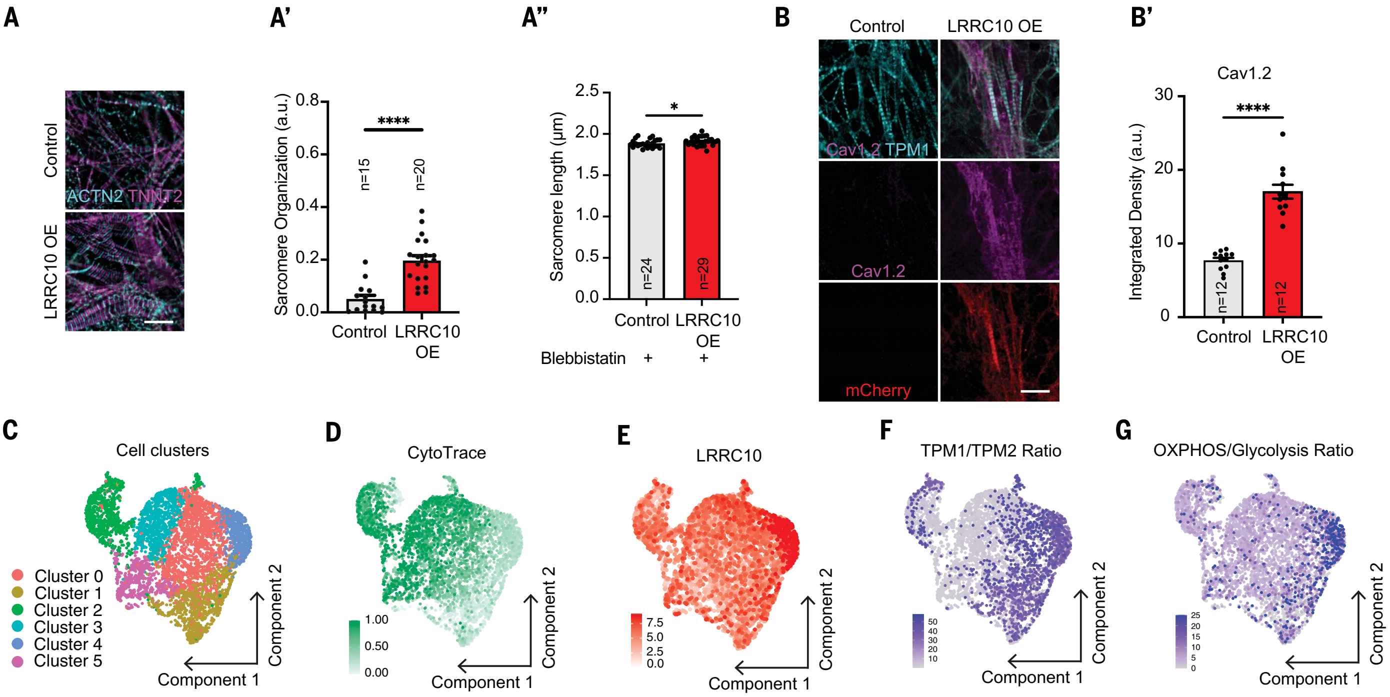


Fig. 3. LRRC10 drives hiPSC-CM maturation. (**A**) Immunostaining with ACTN2 to mark z-disc (cyan) and TNNT2 to mark sarcomeres (magenta) reveals typical disorganized sarcomeres indicating immature iPSC-CMs in the control group and fully assembled sarcomeres in the transfected LRRC10 OE group. Scale bar, 10 μm. (**A'**) Quantification of sarcomere organization in (A) reveals a significant increase after LRRC10 OE compared with control. *t* test was used, and numbers indicate biological replicates. (**A"**) Quantification of sarcomere length in (A) reveals a significant increase after LRRC10 OE compared with control. *t* test was conducted, and numbers indicate biological replicates. (**B**) Immunostaining with mCherry to mark transfected LRRC10 OE, TPM1 to mark z-disc (cyan), and $Ca_v1.2$ to mark LTCC (magenta) reveals highly disorganized dyads, indicating immature hiPSC-CMs in the control group and assembled dyads in the LRRC10 OE group. Scale bar, 10 μm. (**B'**) Quantification of $Ca_v1.2$ signal in (B) reveals a significant increase in fluorescence only in the LRRC10 OE group. *t* test was conducted, and numbers indicate biological replicates. (**C**) UMAP representation of LRRC10 transfected hiPSC-CMs identifies six clusters in the scRNA-seq dataset. (**D**) UMAP representation of scaled CytoTrace scores which indicates the extent of differentiation. A value of 0 (white) is the most differentiated, and a value of 1 (green) is the least differentiated. (**E**) UMAP representation of LRRC10 expression indicates highest expression in cluster 4. (**F**) UMAP representation of TPM1/TPM2 ratio, which represents a sarcomere isoform switch from TPM2 to TPM1, is observed in the maturation axis. (**G**) UMAP representation of OXPHOS/glycolysis ratio, which represents a metabolic switch from expression of glycolytic genes to OXPHOS genes, is observed in the maturation axis.

Lrrc10 expression was significantly reduced in BZ-CMs (fig. S9, C and C'). Similarly, RZ-CMs in a human MI heart showed fully assembled cardiac dyad components located at the z-disc, whereas BZ-CMs located in the recent and old MI regions displayed completely disorganized cardiac dyads (Fig. 4G and fig. S10, A to E'").

Because the cardiac dyad is affected during injury and *lrrc10* may be implicated in its assembly, we next examined how *lrrc10* was driving these dynamics. Transmission electron microscopy on uninjured *lrrc10* sibling hearts revealed that the proteins at the z-disc in *lrrc10*−/− were not as tightly localized to the z-disc as seen in *lrrc10*+/+ (fig. S11, A and A'). We noted that *lrrc10*−/− CMs lacked structures that resembled caveolae (fig. S11B). Next, we examined *lrrc10* in a gain-of-function heart. Complete organization of sarcomeres in BZ-CMs at 7 dpi were observed when disorganized or disassembled sarcomeres should have normally been present (fig. S5, D and D', and fig. S11C). Furthermore, we found relocalization of $Ca_v1.2$ to the dyad region in early regenerating BZ-CMs (7 dpi) (fig. S11C), which would normally be disorganized at this stage in wild-type CMs (Fig. 4D). Thus, Lrrc10 helps to drive the assembly of the cardiac dyad at the z-disc.

Knockdown of *tnnt2a* by antisense morpholino injections impairs sarcomere assembly (*44*). In this case, the sarcomere itself is not assembled properly, but all the sarcomeric components are still present, which provides a condition to test whether Lrrc10 was actively involved in sarcomere reassembly. Although we did not see full rescue of sarcomere assembly in GFP-Lrrc10–overexpressing hearts in the presence of the *tnnt2a* morpholino, we observed increased puncta staining at a higher fluorescence intensity in sarcomeres (fig. S12, A to Diii'). This suggested that the presence of Lrrc10 could direct the cardiac dyad to localize to the z-disc. We tested this directly with fluorescent recovery after photobleaching (FRAP) of Cav1.2-mCherry and Actn2-mCherry in the presence or absence of GFP-Lrrc10. The presence of ectopic GFP-Lrrc10 increased the amount of α-actinin and $Ca_v1.2$ returning to the bleached region of interest (fig. S13, A to C). Thus, Lrrc10 actively drives the reassembly of the cardiac dyad.

Lrrc10 inhibits CM proliferation and can prevent cardiomegaly

CM proliferation is a hallmark of the early regeneration phase, and peak proliferation occurs at 7 dpi (*14*, *26*, *33*), which coincides with the transient peak in *lrrc10* expression (Fig. 2D and fig. S4, A to G). This high *lrrc10* expression negatively correlated with EdU incorporation, a marker of cell proliferation (fig. S14, A and A'). To investigate whether CMs with high *lrrc10* expression had undergone a round of cell cycle activation at an earlier stage, we performed double-pulse chase experiments with an early EdU pulse at 6 dpi followed by a late BrdU pulse at 8 dpi. The high expression of *lrrc10* was associated with EdU+/BrDU− CMs but

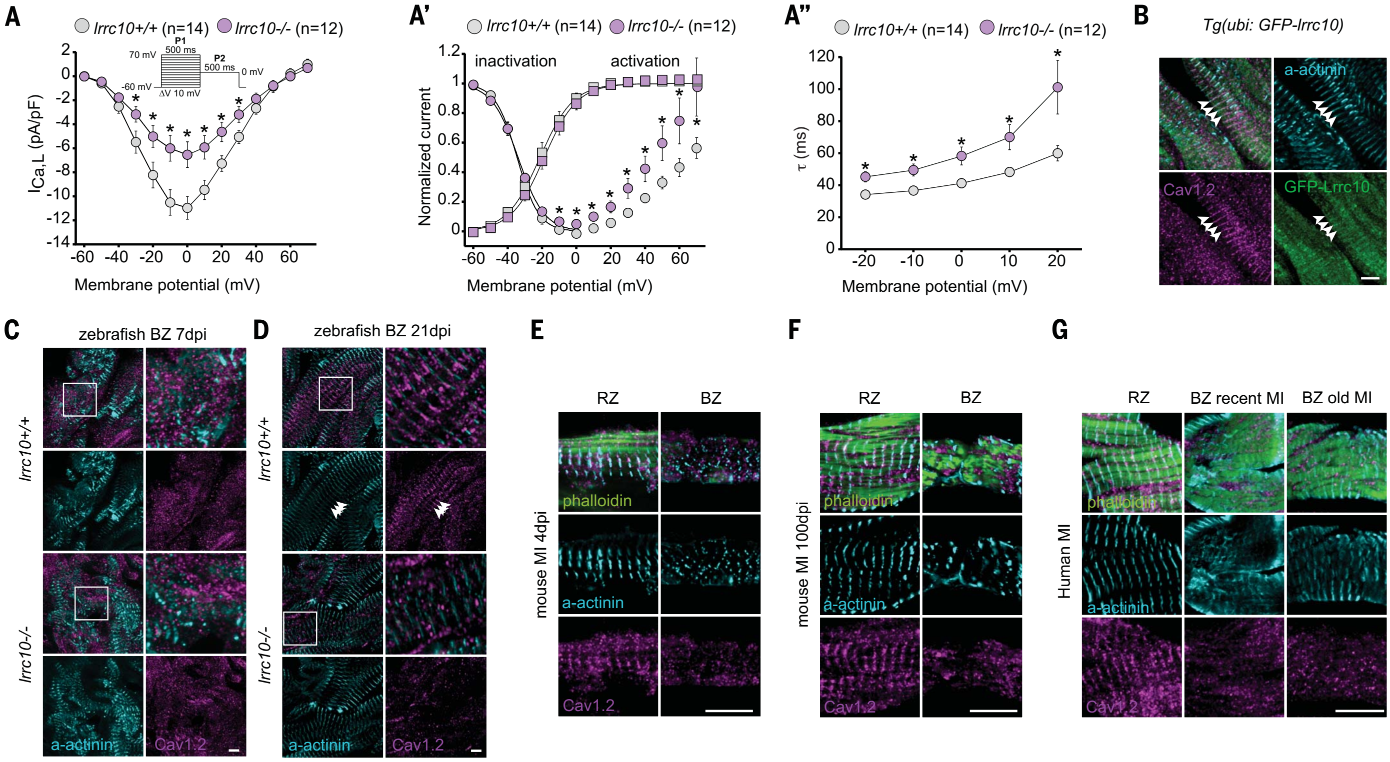


Fig. 4. Lrrc10 is required for cardiac dyad formation during regeneration. (**A**) Patch clamp on isolated uninjured adult CMs reveals a lower current density in *lrrc10−/−* CMs when measuring average current-voltage of $I_{Ca,L}$ current. WT, wild-type. (**A'**) Patch clamp on isolated uninjured adult CMs reveals that *lrrc10−/−* CMs cannot inactivate and have a larger repolarization phase, resulting in more current when measuring voltage dependency of inactivation and activation of $I_{Ca,L}$ current. Activation was not different. Circles represent inactivation, and squares represent activation. (**A''**) Speed (Tau) of $I_{Ca,L}$ inactivation of patch clamped–isolated uninjured adult CMs. (**B**) Antibody staining in *Tg(ubi:GFP-Lrrc10)* uninjured hearts with α-actinin to mark the z-disc (cyan), $Ca_v1.2$ to mark LTCC (magenta), and GFP to mark the localization of the GFP-Lrrc10 fusion protein (green). Arrowheads denote examples of colocalization. Scale bar, 100 μm. (**C**) Antibody staining at 7 dpi in *lrrc10+/+* and *lrrc10−/−* hearts with α-actinin to mark the z-disc (cyan) and $Ca_v1.2$ to mark LTCC (magenta) reveal disorganized dyads. Scale bar, 100 μm. (**D**) Antibody staining at 21 dpi with α-actinin (cyan) and $Ca_v1.2$ (magenta) reveals reassembled dyads (arrowheads) in *lrrc10+/+* but not in *lrrc10−/−* hearts. Scale bar, 100 μm. (**E**) Antibody staining of mouse MI at 4 dpi with phalloidin to mark CMs (green), α-actinin to mark the z-disc (cyan), and $Ca_v1.2$ to mark LTCC (magenta) reveals assembled dyads only in RM-CMs. Images represent four biological replicates. Scale bar, 10 μm. (**F**) Antibody staining of mouse MI at 100 dpi with phalloidin to mark CMs (green), α-actinin to mark the z-disc (cyan), and $Ca_v1.2$ to mark LTCC (magenta) reveals that the BZ-CMs retain disorganized dyads. Images represent four biological replicates. Scale bar, 10 μm. (**G**) Antibody staining in human MI with phalloidin to mark CMs (green), α-actinin to mark the z-disc (cyan), and $Ca_v1.2$ to mark LTCC (magenta) reveals assembled dyads in RM-CMs but not in BZ-CMs at recent and old MI regions. Image represents one replicate. Scale bar, 10 μm.

not with EdU−/BrdU+ CMs (Fig. 5, A and A'), which suggests that *lrrc10* up-regulation is associated with the completion of proliferation and the onset of CM maturation. When examining CM proliferation at the late regeneration stage (21 dpi), there was significant maintenance of proliferation in *lrrc10−/−* BZ-CMs, whereas at the early regeneration phase, *lrrc10−/−* had comparable proliferation rates compared with their wild-type siblings (*37*) (Fig. 5B). Furthermore, Lrrc10 OE caused a decrease in proliferation at the peak CM proliferation time point (7 dpi) (Fig. 5B). Thus, *lrrc10* is a negative regulator of proliferation and an initiator of CM maturity.

We next tested whether the negative correlation of *lrrc10* with proliferation was conserved in mammalian CMs. CM-specific OE of a constitutively active form of the ErbB2 (caErbB2) receptor induces CM dedifferentiation and proliferation resulting in cardiomegaly (*8*). We observed that LRRC10 OE was sufficient to decrease cell cycle activity in caErbB2 CMs (Fig. 5, C and C', and fig. S15, A to C). Moreover, examination of the cardiac dyad showed a significant increase in $Ca_v1.2$ expression but no obvious differences in sarcomere organization upon GFP-LRRC10 OE (fig. S15, D to G). CM proliferation was also examined in hiPSC-CMs that overexpressed LRRC10. LRRC10 OE caused a decrease in EdU incorporation (fig. S15H), pH3 (Fig. 5, D and D'), and a slower growth rate (fig. S15I). Thus, *lrrc10* plays a conserved role in halting CM proliferation across species.

Previous studies in mammals and zebrafish have provided methods to induce CM proliferation by activating specific genes and signaling pathways and have demonstrated positive regenerative outcomes after heart injury (*7*, *8*, *11*, *45*, *46*). A consequence of this uncontrolled CM proliferation, however, is the appearance of large hearts (cardiomegaly). We tested whether the inhibitory effects of Lrrc10 on CM proliferation could prevent cardiomegaly. Continuous 6-week OE of a constitutively active form of vitamin D receptor A (ca-Vdra) (*45*) specifically in CMs resulted in enlarged hearts and visible bulging of the thoracic cavity owing to its excessive size. Notably, ca-Vdra fish coexpressing GFP-Lrrc10 did not show signs of cardiomegaly (Fig. 5, E and E'). To more directly test whether the observed effect was the result of the inhibition of dedifferentiation and proliferation by Lrrc10, we used a vitamin D analog alfacalcidol in adult fish (*45*). Daily injections of alfacalcidol resulted in a significant increase in total ventricle volume in the wild-type group, whereas similar alfacalcidol injections in the GFP-Lrrc10 fish line did not affect total ventricle volume (fig. S16, A

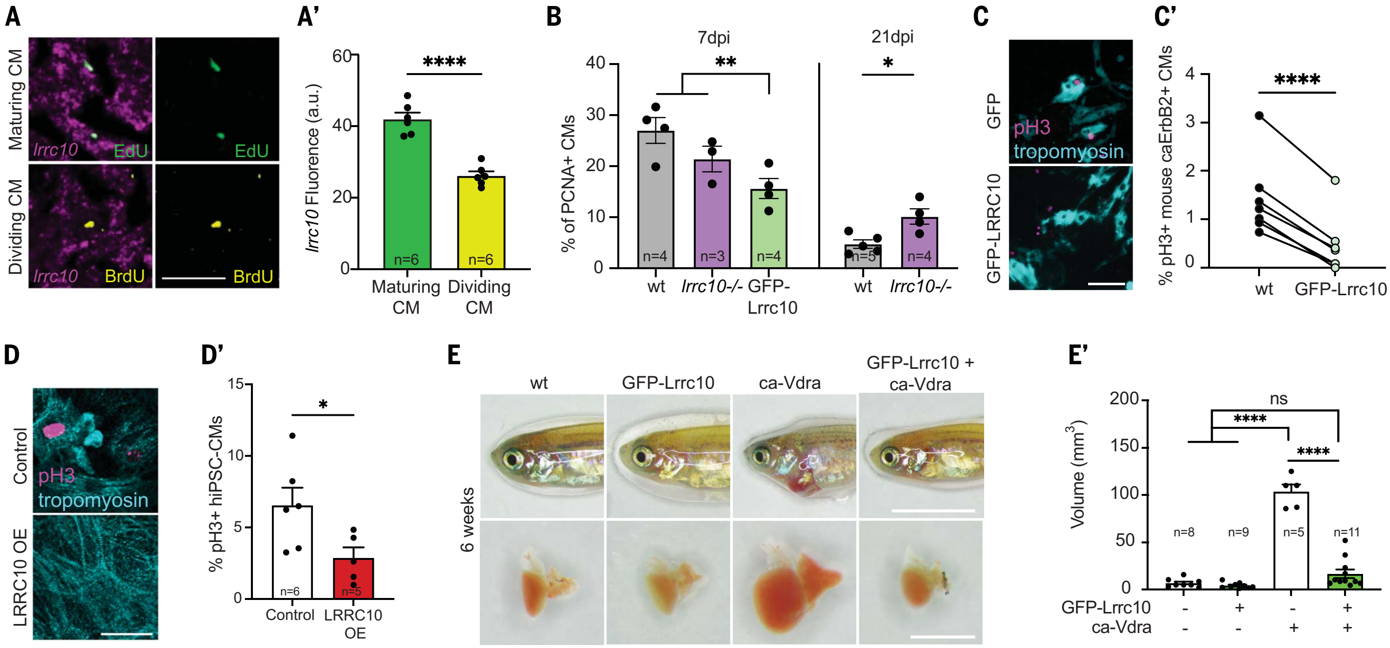


Fig. 5. Lrrc10 inhibits CM proliferation and prevents cardiomegaly. (**A**) RNAscope with *Irrc10* probe (purple) in combination with a maturing CM (EdU+/BrdU–; green) and dividing CM (EdU–/BrdU+; yellow) in 7 dpi BZ-CM. Scale bar, 50 μm. (**A'**) Quantification of *Irrc10* fluorescence signal reveals a negative correlation to proliferation and high *Irrc10* expression. *t* test was conducted, and number indicates biological replicates. (**B**) Cell counting of PCNA$^+$ BZ-CMs reveals similar proliferation level in *Irrc10* siblings at 7 dpi and significantly high level in *Irrc10–/–* BZ-CMs at 21 dpi. Lrrc10 OE causes a significant reduction of proliferation at 7 dpi. *t* test was conducted. (**C**) Antibody staining of mouse caErbB2+ CMs in control GFP only and GFP-LRRC10 OE with proliferation (pH3; magenta) and tropomyosin to mark CMs (cyan). Scale bar, 150 μm. (**C'**) Quantification of proliferation of mouse caErbB2+ CMs after GFP-LRRC10 OE. Paired *t* test was conducted. (**D**) Antibody staining of hiPSC-CMs in control and LRRC10 OE with pH3 to mark proliferation (magenta) and tropomyosin to mark CMs (cyan). Scale bar, 50 μm. (**D'**) Quantification of proliferation in hiPSC-CMs after LRRC10 OE. *t* test was performed, and number represents biological replicates. (**E**) Images of fish and their hearts reveal that wild-type and GFP-Lrrc10 OE fish have comparably sized hearts. ca-Vdra OE shows cardiomegaly, which is rescued when fish coexpress Lrrc10. Scale bar, 2 mm. (**E'**) Quantification of heart volume. One-way ANOVA was conducted, and number represents biological replicates.

to C"). We confirmed that the increase in size was because of proliferation (fig. S16, A to C"). Finally, we tested whether Lrrc10 could prevent cardiomegaly after CM-specific OE of neuregulin1 (Nrg1) (*7*). Control siblings and GFP-Lrrc10 fish receiving phosphate-buffered saline (PBS) displayed normal hearts. Tamoxifen-induced recombined fish that overexpressed Nrg1 displayed an exceptionally large heart. Meanwhile, GFP-Lrrc10 fish that also overexpressed Nrg1 did not show evidence of cardiomegaly (fig. S16, D to E'). Thus, Lrrc10 inhibits CM proliferation, and its OE can robustly prevent cardiomegaly.

Discussion

CM proliferation has been the focus of intense research given its potential to generate new tissue for regeneration. However, although increasing CM numbers is important, the control and subsequent redifferentiation into mature CMs and the integration of these de novo CMs into the native heart is equally essential (*22*). This study provides evidence that rather than being a passive event, CM maturation is an active process. By broadly defining the regenerating CM axis after injury, we propose a mechanism whereby CMs normally use Lrrc10 to bind and recruit components of the sarcomere and LTCC to the cardiac dyad region located at the z-disc. After injury, a downregulation of *lrrc10* permits dedifferentiation of BZ-CMs at a transcriptional and functional level, leading to CM proliferation. Newly generated CMs then up-regulate *lrrc10* to halt the early phase of regeneration and allow CMs to progress through the late regeneration phase, where Lrrc10 aids the reassembly of the cardiac dyad and consequently facilitates the reformation of fully functioning mature CMs. The absence of Lrrc10 leads to incomplete cardiac dyad reassembly and aberrant Ca^{2+} handling and causes CMs to remain in an early regenerative stage, thereby hampering complete regeneration. How exactly Lrrc10 controls Ca^{2+} dynamics and whether regulating Ca^{2+} transients could modulate regeneration need to be further investigated.

Efficient communication between and within individual CMs permits synchronous contraction to produce maximum cardiac output. This remains an unresolved limiting factor for developing cellular-based therapies for heart failure. Specifically, exogeneous cardiac tissue in which CM maturity has not been properly achieved fails to integrate with the preexisting myocardium and exhibits asynchronous contractions; consequently, the long-term contribution to cardiac output is varied (*19*–*23*). This study shows the complex and highly interlinked interaction between sarcomere and Ca^{2+} regulation during CM maturation by using a cardiac regeneration model that naturally generates new CMs. We also found that this link holds true for mammalian CM maturation. We showed that the cardiac dyad is vital for this maturation process, which may provide a potential target to promote CM maturation.

REFERENCES AND NOTES

1. E. Tzahor, K. D. Poss, *Science* **356**, 1035–1039 (2017).
2. K. D. Poss, L. G. Wilson, M. T. Keating, *Science* **298**, 2188–2190 (2002).
3. J. W. Godwin, R. Debuque, E. Salimova, N. A. Rosenthal, *NPJ Regen. Med.* **2**, 22 (2017).
4. E. R. Porrello *et al*., *Science* **331**, 1078–1080 (2011).
5. E. R. Porrello *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **110**, 187–192 (2013).
6. M. Notari *et al*., *Sci. Adv.* **4**, eaao5553 (2018).
7. M. Gemberling, R. Karra, A. L. Dickson, K. D. Poss, *eLife* **4**, e05871 (2015).
8. G. D'Uva *et al*., *Nat. Cell Biol.* **17**, 627–638 (2015).
9. A. Aharonov *et al*., *Nat. Cell Biol.* **22**, 1346–1356 (2020).
10. T. O. Monroe *et al*., *Dev. Cell* **48**, 765–779.e7 (2019).
11. M. Xin *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **110**, 13839–13844 (2013).
12. H. Honkoop *et al*., *eLife* **8**, e50163 (2019).
13. K. Kikuchi *et al*., *Nature* **464**, 601–605 (2010).

14. F. Chablais, J. Veit, G. Rainer, A. Jaźwińska, *BMC Dev. Biol.* **11**, 21 (2011).
15. C. C. Wu *et al*., *Dev. Cell* **36**, 36–49 (2016).
16. R. Fukuda *et al*., *EMBO Rep.* **21**, e49752 (2020).
17. A. Beisaw *et al*., *Circ. Res.* **126**, 1760–1778 (2020).
18. A. Lepilina *et al*., *Cell* **127**, 607–619 (2006).
19. Y.-W. Liu *et al*., *Nat. Biotechnol.* **36**, 597–605 (2018).
20. J. J. Chong *et al*., *Nature* **510**, 273–277 (2014).
21. Y. Shiba *et al*., *Nature* **538**, 388–391 (2016).
22. A. Shakked *et al*., *Nat. Cardiovasc. Res.* **2**, 383–398 (2023).
23. K. Gabisonia *et al*., *Nature* **569**, 418–422 (2019).
24. C. Jopling *et al*., *Nature* **464**, 606–609 (2010).
25. V. Gupta, K. D. Poss, *Nature* **484**, 479–484 (2012).
26. A. Bertozzi *et al*., *Dev. Biol.* **471**, 106–118 (2021).
27. D. A. Eisner, J. L. Caldwell, K. Kistamás, A. W. Trafford, *Circ. Res.* **121**, 181–195 (2017).
28. R. E. Ahmed, T. Anzai, N. Chanthra, H. Uosaki, *Front. Cell Dev. Biol.* **8**, 178–178 (2020).
29. T. W. Chen *et al*., *Nature* **499**, 295–300 (2013).
30. H. Sánchez-Iranzo *et al*., *Nat. Commun.* **9**, 428 (2018).
31. G. S. Gulati *et al*., *Science* **367**, 405–411 (2020).
32. X. Qiu *et al*., *Nat. Methods* **14**, 979–982 (2017).
33. J. M. González-Rosa, V. Martín, M. Peralta, M. Torres, N. Mercader, *Development* **138**, 1663–1674 (2011).
34. K. Schnabel, C. C. Wu, T. Kurth, G. Weidinger, *PLOS ONE* **6**, e18503 (2011).
35. M. T. Woon *et al*., *J. Am. Heart Assoc.* **7**, e006428 (2018).
36. W. Feng *et al*., *Circulation* **141**, 940–942 (2020).
37. W. T. Stockdale *et al*., *Cell Rep.* **25**, 1997–2007.e7 (2018).
38. X. K. Qu *et al*., *Mol. Med. Rep.* **12**, 3718–3724 (2015).
39. M. J. Brody *et al*., *PLOS ONE* **7**, e51621 (2012).
40. E. Giacomelli *et al*., *Cell Stem Cell* **26**, 862–879.e11 (2020).
41. F. Hofmann, V. Flockerzi, S. Kahl, J. W. Wegener, *Physiol. Rev.* **94**, 303–326 (2014).
42. M. J. Brody *et al*., *Am. J. Physiol. Heart Circ. Physiol.* **310**, H269–H278 (2016).
43. K. H. Kim, T. G. Kim, B. K. Micales, G. E. Lyons, Y. Lee, *Dev. Dyn.* **236**, 2225–2234 (2007).
44. W. Huang, R. Zhang, X. Xu, *Dev. Biol.* **331**, 237–249 (2009).
45. Y. Han *et al*., *Dev. Cell* **48**, 853–863.e5 (2019).
46. M. Ogawa *et al*., *Science* **372**, 201–205 (2021).

ACKNOWLEDGMENTS

We thank D. Ward-van Oostwaard (LUMC) for assistance with hiPSC-CM experiments; B. de Jonge (AMC) for assistance in zebrafish isolation for patch-clamp experiments; D. Verteeg (Hubrecht Institute) and E. van Rooij (Hubrecht Institute) for performing the original mouse MI procedures; D. Y. R. Stainier (MPI Bad Nauheim) and M. Mommersteeg (University of Oxford) for sending zebrafish lines; K. D. Poss (Duke University) for sending plasmids; J. Stein (LUMC) for providing the SoTa tool for sarcomere analysis; M. Dias Brescia (LUMC) and K. Barnes (Sartorius) for input into the real-time imaging and Sartorius for use of the Incucyte Live-Cell Analysis System; S. L. Kloet, E. de Meijer, and R. Menafra (Leiden Genome Technology Center) for help with 10X Genomics scRNA-seq (advice, library preparation, primary data alignment, and quality control); and F. Tessadori (Hubrecht Institute) for critical review of the manuscript. **Funding:** This study received support from an EMBO Long Term Fellowship ALTF1129-2015 (P.D.N.); an HFSPO Fellowship (LT001404/2017-L) (P.D.N.); NWO-ZonMW Veni grant (016.186.017-3) (P.D.N.); European Union's Horizon 2020 Research and Innovation Programme under the Marie Sklodowska-Curie action (MSCA-IF 838985 SiGNATURE) (G.C.); Netherlands Organ-on-Chip Initiative, an NWO Gravitation project funded by the Ministry of Education, Culture and Science of the government of the Netherlands (024.003.001) (C.L.M.); the European Research Council (ERC-CoG Mini-HEART no. 101001746) (M.Be.); the Novo Nordisk Foundation Center for Stem Cell Medicine supported by Novo Nordisk Foundation grants (NNF21CC0073729) (C.L.M.); the European Research Council (ERC-StG IniReg 716894) (K.B.); the Netherlands Cardiovascular Research Initiative, an initiative with support of the Dutch Heart Foundation and Hartekind (CVON2019-002 OUTREACH) (J.B.); and NWO-ZonMW Open competition grant (CONTRACT 09120012010018) (J.B.). **Author contributions:** Conceptualization: P.D.N. and J.B. Investigation: P.D.N., G.C., I.G., A.O.V., H.H., M.Bo., D.E.M.d.B., T.K., A.V., G.E.M.L., A.S., J.M., A.A.M., S.C., and T.P.d.B. Funding acquisition: P.D.N., K.B., E.T., C.L.M., M.Be., and J.B. Supervision: P.D.N., K.B., E.T., C.L.M., M.Be., and J.B. Writing – original draft: P.D.N. and J.B. Writing – review & editing: all authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. Sequencing datasets can be accessed in GSE224156. The hiPSC line LUMC0020iCTRL-06 (https://hpscreg.eu/cell-line/LUMCi028-A) is available from M.Be. (Anatomy and Embryology Department) or C. Freund (iPSC core facility) under a material transfer agreement with Leiden University Medical Center. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abo6718
Materials and Methods
Figs. S1 to S16
Table S1
References (*47*–*72*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 24 February 2022; resubmitted 26 January 2023
Accepted 20 April 2023
10.1126/science.abo6718

TECHNICAL COMMENT

POLICY FORUM

Technical comment on "Fairness considerations in global mitigation investments"

Gregor Semieniuk[1]*, Jayati Ghosh[1], Nancy Folbre[1]

In an otherwise excellent analysis of fair regional shares of global mitigation investments, Pachauri *et al.* (Policy Forum, 9 December 2022, p. 1057) dramatically overestimate developing countries' 'capability' to invest by estimating GDP using purchasing power parity exchange rates. Since internationally sourced investment goods must be paid for at market exchange rates, capability-based interregional finance flows should be vastly larger.

Pachauri *et al.* (*1*) have calculated fair regional shares of financing global climate change mitigation investments through 2030 using operational versions of 'responsibility', 'capability', and 'needs'. The authors find that these fairness considerations imply scaled up interregional investment flows. Higher capability, which is based on affluence (per capita GDP or capital stock), translates into higher fair investment shares. We show that the authors' use of exchange rates at purchasing power parity (PPP) to calculate 'capability' dramatically overestimates the capability of developing countries to finance investments.

PPPs are artificially constructed exchange rates that are supposed to reflect the ability of a currency to purchase goods and services in its country of issuance or use compared with the US-dollar in the United States. Since domestic production of goods and services in lower income countries tends to be cheaper on account of lower wages and cheaper domestically produced intermediate inputs, these goods and services' prices are lower than in a rich country, after converting currency at market exchange rates. Therefore, PPP exchange rates tend to appreciate the currency of lower income countries. As a result, GDP per capita converted to US-dollars at observed market exchange rates (MER) is vastly lower for most countries. For lower-middle income countries such as India or Indonesia MER GDP is only around one-third of that calculated at PPP; for upper-middle income countries such as China or Brazil the figure is around half. GDP per capita inflation when applying PPP exchange rates for developing countries is so pervasive that it has its own name: the Penn effect, after the university where these synthetic exchange rates were first systematically estimated (*2*, *3*).

Since the value of PPP exchange rates depends on many assumptions and changes with each new estimation exercise (the International Comparison Project), there is a substantial literature critiquing the specific values or the use of PPP exchange rates as well as adjusting the methodology in response (*4*–*8*). Such critiques, e.g., that PPPs tend to overestimate real incomes in lower income countries (*9*), are a serious problem for using PPP GDP to measure a capability to invest.

Our critique here is more specific. MER are simply the more appropriate indicators of purchasing power for internationally traded goods and services since these are the exchange rates

[1]Political Economy Research Institute and Department of Economics, University of Massachusetts Amherst, Amherst, MA 01003, USA.
*Corresponding author. Email: gsemieniuk@umass.edu

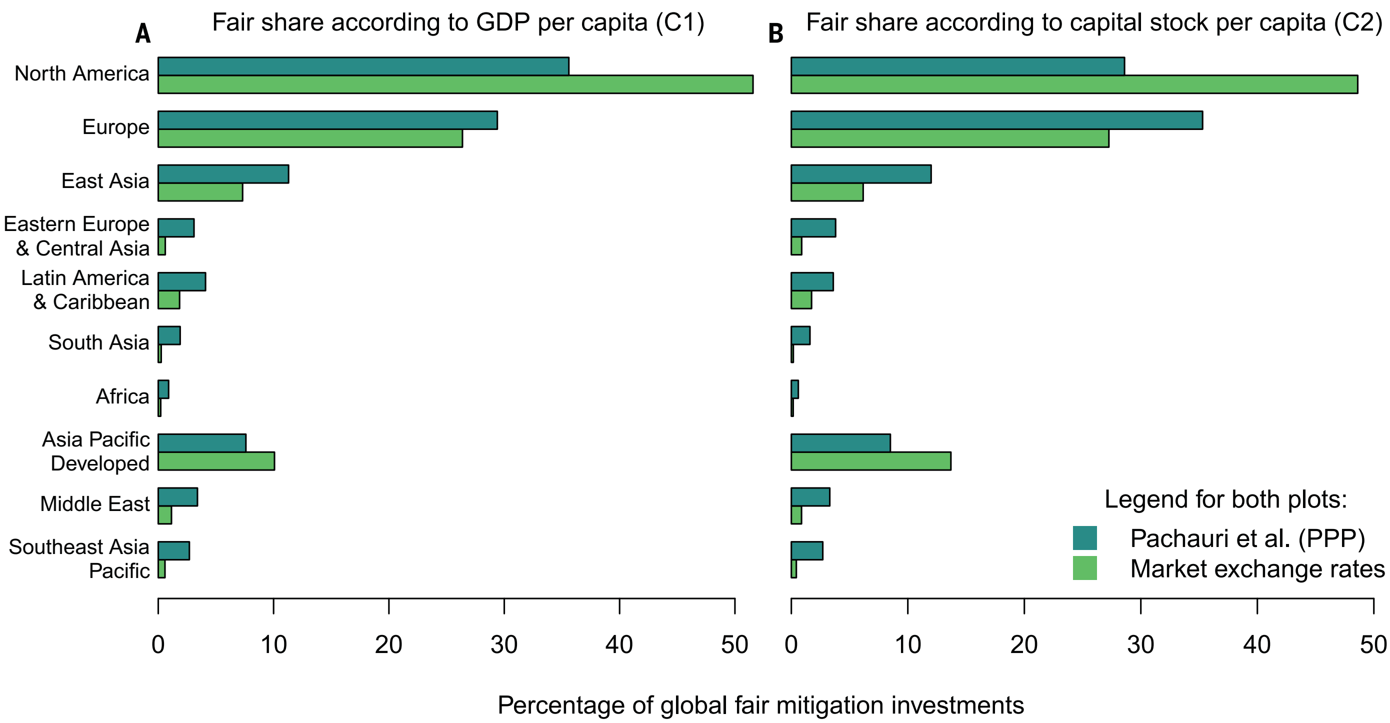


Fig. 1. Recalculated fair capability shares in context of the existing results. Fair share of investments according to PPP-based capability reported in Pachauri *et al.* (*1*) (dark green) and recalculated MER-based capability (light green) for each of 10 regions of the world. (**A**) shows results for capability according to GDP per capita [corresponding to C1in figure 2 from (*1*)], (**B**) for capability according to capital stock per capita [corresponding to C2 in figure 2 from (*1*)]. See (*14*) for details on data, code, and results.

at which international transactions actually occur. MER are key for financing climate change mitigation investments. Most of the essential plant and equipment for low-carbon energy supply and end-use technologies is sourced internationally and must be paid for in US-dollars obtained at MER (*10*–*12*). For instance, 90% of all wind turbines are manufactured in only 15 countries, 90% of silica-based solar PV modules are manufactured in East Asia, electric vehicle batteries are exclusively manufactured in East Asia, North America, and Europe, and electrolyzers are manufactured exclusively in China, Europe, North America, Japan, and India (*13*). This means that the capability to pay for mitigation investment through 2030 must be assessed in terms of GDP calculated at market exchange rates.

Recalculating the capability to invest using GDP at market exchange rates leads to a substantial readjustment of the authors' results. We first replicated the capability results with the PPP approach of Pachauri *et al.* (*1*) following the instructions and results reported in their supplementary materials, and using publicly available data from the World Bank (GDP) and Penn World Tables (capital stock) [complete results, data and figure code available at (*14*)]. Then we replaced all of their PPP-based conversions with a MER-based conversion from the same sources. Figure 1 shows in dark green bars the results for capability that Pachauri *et al.* obtained, and in light green bars the results using the MER approach. Comparing these bars shows that in all regions but Asia Pacific Developed and North America, the fair share of investment falls when calculated at MER GDP. For instance, Southeast Asia Pacific's share drops from 2.7% to 0.6% under MER-recalculation of capability according to GDP per capita, South Asia's share from 1.9% to 0.3% and Latin America's share from 4.1% to 1.9%. By contrast, North America should pay for 52% instead of 36% of all investments. All regions outside of the three richest ones (North America, Asia Pacific Developed, and Europe) should together only pay for 12.0% instead of 27.4%. The result is analogous for the capability to invest based on capital stock per capita, whose value the authors also calculate using PPP exchange rates, and where the developing regions' joint share drops from 27.6% to 10.4% with MER recalculation.

Discussions of mitigation burden-sharing need to take seriously the major implications of the difference between PPP and MER exchange rates for trade-related considerations in climate mitigation finance modeling. In terms of market exchange rate, the capability to finance investments of the richest regions suggests nearly 90% of all investments until 2030 should be financed by them. If done well, such cross-border investments could also vastly lower the financing costs of the global energy transition (*15*).

REFERENCES AND NOTES

1. S. Pachauri *et al.*, *Science* **378**, 1057–1059 (2022).
2. P. A. Samuelson, *Rev. Int. Econ.* **2**, 201–226 (1994).
3. B. A. Heston, R. Summers, *Am. Econ. Rev.* **86**, 20–24 (1996).
4. T. Pogge, S. G. Reddy, *SSRN* **41**, 2241–2247 (2006).
5. S. Anand, P. Segal, *J. Econ. Lit.* **46**, 57–94 (2008).
6. S. Johnson, W. Larson, C. Papageorgiou, A. Subramanian, *J. Monet. Econ.* **60**, 255–274 (2013).
7. A. Deaton, B. Aten, *Am. Econ. J. Macroecon.* **9**, 243–264 (2017).
8. A. Deaton, P. Schreyer, *Rev. Income Wealth* **68**, 1–15 (2022).
9. J. Ghosh, *Econ. Labour Relat. Rev.* **29**, 24–37 (2023).
10. K. Surana, C. Doblinger, L. D. Anadon, N. Hultman, *Nat. Energy* **5**, 811–821 (2020).
11. P. Mealy, A. Teytelboym, *Res. Policy* **51**, 103948 (2022).
12. G. Semieniuk, L. Taylor, A. Rezai, D. K. Foley, *Nat. Clim. Chang.* **11**, 313–318 (2021).
13. IEA, *Energy Technology Perspectives* (International Energy Agency, 2023).
14. G. Semieniuk, J. Ghosh, N. Folbre, Figure data and code used in Technical comment on "Fairness considerations in global mitigation investments", Zenodo (2023).
15. D. Matthäus, M. Mehling, *Joule* **4**, 2627–2645 (2020).

ACKNOWLEDGMENTS

Competing interests: Authors declare that they have no competing interests. **Data and Materials Availability:** Data and code used in the analysis are available from (*14*). G.S. is affiliated with the University of Sussex, University College London and SOAS University of London.

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adg5893
Data S1

Submitted 22 January 2023; accepted 12 April 2023
10.1126/science.adg5893

TECHNICAL RESPONSE

POLICY FORUM

Response to comment on “Fairness considerations in global mitigation investments”

Shonali Pachauri[1]*, Setu Pelz[1], Christoph Bertram[2], Narasimha D. Rao[1,3], Keywan Riahi[1]

We welcome the analysis of Semieniuk *et al*. (*1*) as an additional sensitivity to illustrate a more extreme distribution of regional contributions to climate mitigation investments that supports our main conclusion regarding the North-South divide in mitigation investment capabilities. In response to Semieniuk *et al*. we would like to first point out that, in defining the required global mitigation investments for the 2020 to 2030 period, our study relies on the estimates in the sixth assessment report (AR6) of the Intergovernmental Panel on Climate Change (IPCC) WGIII (*2*). These are based on diverse sources and underlying models that to varying degrees reflect regional differences in technology costs and consider both purchasing power parity (PPP) and market exchange rates (MERs). We use these IPCC estimates as a starting point and focus entirely on the question of how much of the needed regional investments, given different fairness considerations, should be financed from sources within a region.

We agree with Semieniuk *et al.* that applying MER would indeed result in higher differences between countries' relative capabilities to finance low-carbon investment needs. We argue, however, that using PPP is an approximation that is more legitimate in the context of our analysis. This is primarily because it is established practice in the literature to use PPP exchange rates for comparing the relative capabilities of different countries (*3*). For instance, the European Commission uses the GDP in PPPs of member states to set financial transfers between countries when allocating structural funds (*4*). PPP exchange rates can therefore also be considered an appropriate base for calculating countries' capacity to finance investments in other regions. The opportunity costs of international flows to finance investments elsewhere would be investment within a given region, which, by definition, should be valued at prices in the region. PPP exchange rates are thus more appropriate in reflecting these opportunity costs as they approximate how much of the currency of one country would be needed to buy a similar basket of goods and services in another country. MERs do not capture cost of living differences or reflect the different abilities to pay for goods and services in different countries.

Second, we think it is questionable whether the share of international trade in climate mitigation investments is as substantial as stated by Semieniuk *et al.* This is because a considerable fraction of mitigation investments, particularly in LMICs, will arise on the demand side, e.g., building retrofits and efficient new buildings, mobility infrastructure, the installation of PV panels etc., which are dominantly local investments requiring local labor and materials. Even considering essential powerplant and equipment for low-carbon energy supply, the local components, including construction costs, labor, etc. are considerable. This is evident also, for instance, when one compares the levelized cost of energy (LCOE) of solar photovoltaics (SPV) across countries, which differs widely even though the PV cells themselves are usually imported from a limited number of suppliers (*5*).

Finally, the use of PPP exchange rates is also consistent with the wider literature on international income inequality and comparative capabilities. PPP exchange rates provide a standardized measure for comparing investments across countries with different economic structures and levels of development. In contrast to MER, PPP correct for differences in the cost of living and production factors across countries. This is critical for understanding the capacity of different regions or nations for financing mitigation investments and the related opportunity costs of not doing so. Thus, although MERs may provide a clearer picture of the actual costs of internationally traded goods and services, we feel that MERs are less appropriate as measures of comparative capability, which is necessary for our analysis, as much of the mitigation investments needed in LMICs are likely to be dominated by local investments.

[1]International Institute for Applied Systems Analysis, Laxenburg, Austria. [2]Potsdam Institute for Climate Impact Research, Potsdam, Germany. [3]Yale School of the Environment, New Haven, CT, USA.
***Corresponding author. Email: pachauri@iiasa.ac.at**

REFERENCES AND NOTES

1. G. Semieniuk, J. Ghosh, N. Folbre, Comment on “Fairness considerations in global mitigation investments” (Science, 2023); 10.1126/science.ade6084.
1. S. Pachauri *et al*., *Science* **378**, 1057–1059 (2022).
2. IPCC, *Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, 2022).
3. W. Nordhaus, *Energy Econ.* **29**, 349–372 (2007).
4. Eurostat European Commission, OECD, *Eurostat-OECD Methodological Manual on Purchasing Power Parities*, (OECD, 2012).
5. The International Renewable Energy Agency (IRENA), “Renewable Power Generation Costs in 2019” (IRENA, 2020).

10.1126/science.adh1463

WORKING LIFE

By **Claudia Fracchiolla**

The power of outreach

I was crying alone on the emergency exit stairs outside my lab when a professor I worked with as a teaching assistant came by. He kindly checked in about how I was doing and went on to tell me about a project he had found rewarding: sharing physics demos with schoolchildren. That resonated with me. Earlier in my career I had always looked for opportunities to engage the community through outreach. But since I started my Ph.D., seeking out public engagement opportunities had seemed an unwise distraction from my classes and research. Now, I wondered if I should reconsider.

I was nearing my breaking point. I had come to the United States for my doctorate after completing my undergraduate degree in my home country of Venezuela and then continuing my studies in Brazil and Europe. Here, for the first time, I was experiencing discrimination for being Latina. Some professors said they could not understand my English; I couldn't attend office hours without a classmate to translate. And I was struggling with the department's preferred teaching and assessment style in my classes, which was very different from what I was used to.



"Doing public outreach ... reminded me why I wanted to be a scientist in the first place."

During that stairwell discussion with the professor, I shared some of my previous experience engaging in outreach, and he invited me to join the program he ran, putting together physics demos to present at schools and other venues. The offer lifted my spirits, but I still wasn't sure. Would it take away valuable time from my research and studies? Would my professors disapprove?

But as I reflected on it over the next few weeks, I realized I couldn't afford not to explore this opportunity. In the past, doing public outreach left me reinvigorated to return to my "regular" work and reminded me why I wanted to be a scientist in the first place. Plus, talking about outreach with the professor was the first time I had felt truly seen and valued since moving to the United States. I desperately needed more of that to be able to carry on.

From my first day on the project, I felt welcomed and included. As I started to put together demos to present at a children's hospital, I began to feel more energized and driven. I did get a cautionary comment from a professor reminding me to focus on my research—but I was too excited for that to get me down. We produced displays that let children at the hospital and their families create beautiful patterns of water vibrating to the rhythm of their hands and attempt to stand on a board supported by paper tubes. Watching their enthusiasm reminded me why I fell in love with physics.

That experience solidified my love for public engagement. However, because it was not laboratory research, some colleagues and professors questioned its relevance to my physics training. Ultimately, their skepticism made me feel I had to choose between pursuing "hard," "serious" physics research or doing "soft" engagement work. I went on to complete a Ph.D. in science education and pursue a career in public engagement. I love what I do, but it still hurts that I felt I had to leave traditional research to do it.

In physics education research studies I led as a postdoc, I learned that many scientists—particularly those from underrepresented groups—find public engagement helps sustain them through tough times in their careers, just as it did for me. One student shared with me that facilitating the children's hospital project inspired her to persist in physics; being recognized as an expert by the public helped her overcome a sense that she didn't belong in academia. I had felt the same.

In a more supportive environment, engaging in outreach can help scientists feel valued and inspire them to stay the course amid difficulties. Unfortunately, many of us face the perception that the more time a scientist spends popularizing science, the less scientifically accomplished they are. That is far from true—and it drives far too many of us out of traditional research. ■

Claudia Fracchiolla is the head of public engagement for the American Physical Society. Send your career story to SciCareerEditor@aaas.org.